

13 May 1982

Will there be war in space?

The sheer scale of the Soviet space programme is awesome. Has the time come when the Soviet government must say what it hopes to achieve?

In June the name of Jean-Loup Chrétien will become a household word in the Soviet Union and France. Chrétien will be the first Western astronaut to be whooshed into space by a Soviet rocket and to ride in a *Soyuz* spacecraft, visit a Soviet space station briefly and, if all goes well, return to Earth. He will join the list of Polish, Czech, East German, even Cuban, cosmonauts whom the Soviet Union has made guests in space, and who have been given a first-hand look at the civilian side of that technology about which the Soviet Union is most sensitive and proud: manned space flight. At about the same time, in Washington, where the Soviet space programme generally draws a big yawn, the President will be in the final stages of a critical decision on whether to proceed with a US manned space station to serve the United States and Western Europe.

The Soviet space programme has received little attention in Europe or the United States. Yet its objectives and momentum will influence East-West relations and the arms race in the 1980s and beyond. True, the Western media are delighted to talk up "Star Wars" fantasies, the US shuttle, Ariane or Soviet space weapons. But they have overlooked the main substance of the Soviet programme, lucidly outlined in a recent book by NASA watcher of the Soviet skies, James E. Oberg*. Since the early 1970s, the Soviet manned space programme, which had been in a shambles after the death of its early architect Sergey Korolev in 1966, has concentrated on a single goal: *Kosmograd*. This is the code word for "Space Colony Number 1, made in the Soviet Union". In an assessment shared by many observers of the Soviet manned space programme, Oberg describes how the Soviet Union is amassing the building blocks of this capability.

First there was *Soyuz*, the free-flying craft in which cosmonauts journey up and down from space and make (brief?) orbital stays for scientific work or military reconnaissance. These flights started in 1967 and numbered 40 by May 1981. Then came the *Salyut* space stations, the first one launched in 1971. These have been orbiting space homes for repeated and successively longer visits by cosmonauts. The world's record of 185 days was reached aboard *Salyut 6* in 1980. One of the astronauts who performed the feat had just spent another 175 days in space, and so has had nearly a year of weightlessness. Finally came *Progress*, the unmanned robot freighter tanker craft described in Moscow as a "reliable cargo barge to ensure year-long operation by orbital stations". *Progress* craft bring lonely cosmonauts fuel, food, fresh water, medicines, horseradish, newspapers, mail and other things. By now there is extensive experience in sending up these vehicles, docking the *Soyuz* and *Progress* craft with the *Salyut*, moving them from one *Salyut* port to another to admit a second visiting craft and making repairs in space, including emergency repairs to the exterior of the craft.

So engineering experience is accumulating, but *Kosmograd* also requires mastery of space biology and medicine as well as experience with prolonged human confinement. The goal of sustaining life in space for a year or more is within sight. Although the large green cucumber that the crew on one *Salyut* showed to astonished mission controllers on television proved to be a joke, with practice the Soviets have managed to get seeds growing in space. Space gardens could provide 20 to 40 per cent of the bulk food consumed in orbit, but it remains a mystery why space-grown plants have not yet grown flowers or seeds. Water aboard

the *Salyut* is now 50 per cent recycled, and the addition of a urine still could achieve 90 per cent recycling. Banks of *Chlorella* algae have been used to absorb carbon dioxide and give off fresh oxygen: 80 per cent recycling has been achieved on this front. But things have not always gone well. Cosmonauts set to achieve a new endurance record aboard *Salyut 4* pleaded early on to come home because a green mould was spreading all over the cabin. Mission control ignored these pleas until the endurance record had been broken (61 days).

The objective of all this is a permanent orbiting station. *Salyut*, housing two, will be succeeded soon by a larger spaceship housing twelve. According to Oberg, these craft could even be sent to the Moon with the help of the huge proton booster (whose shadowy cousin, the superbooster, apparently failed so often in the late 1960s that it prevented a photofinish in the race to the Moon against the United States).

But such a feat will, on past experience, require a high price in human life. Cosmonauts may be heroes in the Soviet Union, but few seem to live to enjoy it. Recovery teams have often arrived at a returned capsule dangling by its parachute lines in a tree or sitting in some swamp to find the entire crew dead. There was even an unauthorized space-walk in which the cosmonaut's life support cord was unattached and when his life was saved only by the quick thinking of his colleague. Yuri Gagarin, whose 108-minute flight in 1961 amazed the world, became a national hero but he was fated to eulogize comrade after comrade at one funeral after another. (His own life was thrown away cheaply in an aircraft crash.) Even Korolev's death on an operating table, which cost the Soviet space programme its single most important figure, seems to have been unnecessary. There have also been close shaves with death, as when the cosmonauts who missed their landing area by 2,000 miles had to spend their first night back on Earth in the Siberian snows, hiding in their spacecraft from the wolves that snapped outside.

What is the underlying purpose of all this activity? There are obvious civil benefits, in telecommunications for example. And by now, both the United States and the Soviet Union take for granted the stabilizing effects of each other's use of satellites for surveillance. Even so, there is growing concern in the United States at what seems to be the huge scale of the military part of the Soviet space programme. Oberg says there are two "star cities", one for cosmonauts and their guests and the other for the military. A recent study by Marcia Smith of the Congressional Research Service concludes that as much as 70 per cent of the Soviet effort may be exclusively military, while nobody has yet contradicted the assertion by Richard de Lauer, chief scientist at the Pentagon, that spaceborne Soviet lasers could threaten US spacecraft as early as 1983. Technical considerations make the speculation unconvincing; for the time being conventional rocket-driven explosives are probably a better bet. But as in other aspects of the relationship between the United States and the Soviet Union, policy may be determined by one state's perceptions of what the other is about. The fact that the Soviet Union successfully launched 98 satellites last year, compared with a mere 18 from the United States, is certain to carry weight with the White House.

What will be the outcome? Since the United States and the

*Oberg, J.E. *Red Star in Orbit* (Harrap, London and Random House, New York).



Soviet Union ratified the Outer Space Treaty in 1967, they have been bound not to put nuclear weapons or other "weapons of mass destruction" into orbits about the Earth. The treaty also requires them not to establish military bases on celestial bodies. Would that injunction apply to a space station occupied exclusively by the military? Or to a satellite equipped to knock out other people's satellites but which contained no people, only instruments? The Soviet Union should come clean about the objectives of its massive programme. The terrestrial arms race is uncomfortable enough. The prospect of a race to put flesh on the bones of science fiction is alarming. Meanwhile, the odds are shortening that President Ronald Reagan, due to greet the next return of the shuttle *Columbia* at Edwards Air Force Base in California on 2 July, will not merely proclaim that the shuttle is a going concern but that the United States will embark on the next logical step, the building of a space station. As Kennedy did before him, Reagan may find that the Soviet Union's activities justify an expanded programme in the United States.

Academic suicide

The University of London now has a plan for its future; but will it survive that long?

The University of London seems well on the way to making an even greater hash of its affairs than seemed likely a few months ago (see *Nature* 14 January, p.86). For the past year, it has been clear that the consequences of the British government's reduction of support for universities, the limits imposed by the University Grants Committee on the number of students to be allowed at individual universities and the effect on the income of universities of frightening away students from overseas would be especially damaging for the largest British university and, because of its federal constitution, administratively the most complicated. But none of that can excuse the way in which the university has let time slip through its fingers until last week when, with the deadline approaching for deciding how it should finance itself next year, the university's senate was jockeyed into accepting a plan for the future whose analytical foundations are, to say the least of it, shaky and whose academic implications are unknown. The only virtue in the way in which the university has managed its affairs is that this sorry tale may serve as a warning to other academic institutions.

What has happened in the University of London is a kind of pantomime. Two years ago, the then vice-chancellor, seeing the way the wind was blowing, set up a Committee on Academic Organization under an independent chairman. The plan was to recommend a strategy for the university as a whole. Halfway through this undertaking, however, a new vice-chancellor gave the committee narrower terms of reference and set up six other committees, again with independent chairmen, to suggest how the pattern of teaching in different fields of study should be reorganized. (One committee went off on the wrong tack, preferring terms of reference other than those given to it; mercifully, with a different chairman; a set of recommendations has been cobbled together just in time.) Inevitably, these subject committees have produced proposals for moving the teaching of various subjects from one place to another in the federated university which are inevitably in conflict with each other, so yet another committee — called the Joint Planning Committee — was asked to make sense of them and of the volume of comment and special pleading with which the university was deluged. That committee concludes, in its report, that the recommendations of the six subject committees were often based "on inadequate information" (ill-informed) or "cannot be justified in the light of the facts" (or in simple language are wrong). Searching, nevertheless, for some straw at which to clutch, the committee seems to have seized on one of the recurrent themes in the recommendations of the subsidiary committees — physics, or classics, or economics, should be taught at fewer sites. So why not simply concentrate the whole of the university, now spread

between a dozen non-medical colleges, on a smaller number of places, say five or six? This is the proposal that the senate of the university found itself having to accept last week.

The most curious feature of the affair is that this radical decision has been reached with hardly any discussion in the university as a whole of the academic implications, let alone the implications for those who work in and for the university. And what is now intended will create further internal contradictions within what is already a sufficiently cumbersome academic machine. The university grew to its present size like Topsy, by the enthusiastic accretion of almost any college, large or small, willing to join the fold. As a result, it now consists of an exceedingly inhomogeneous collection of establishments, some large and some small, some academically excellent, some less so. Two are for practical purposes specialized in the scientific fields. One (Birkbeck College) specializes in part-time students. One small college is not in London at all but in Surrey, while there are agricultural and veterinary colleges in Kent and Hertfordshire respectively. The tidy administration of such a motley collection of establishments is acknowledged to be impossible. Since the late 1960s, the university has chosen to make a virtue out of diversity, letting its separate parts run themselves without much reference to the centre. Now, by a simple show of hands, that policy has been turned on its head. The committee whose recommendations will shape the future of the university pays the usual lip-service to diversity before recommending that it should be done away with.

The university's problem is uniquely difficult — which is all the more reason why it should have been dealt with intelligently. Given that the pressure on the university's finances is likely to continue more or less indefinitely, a measure of concentration is clearly in the long run prudent. Some of the London colleges, seeing the way the wind is blowing, are already making mergers among themselves. The merger (or disappearance) of the rest will now no doubt be forced by the decisions taken centrally by the university on the distribution of funds among its several parts in the years immediately ahead. In the process, it is inevitable that much damage will be done. Many of the shotgun marriages the university is now committed to arranging will be marriages of incompatibles. Undergraduate teaching will become more specialized, and more uniform. The pattern of research will be changed in an arbitrary fashion. And the end result will be a federation of colleges so much more like each other that the advantages of federation will melt away. If the arrangements now agreed are pushed through, it cannot be long before the various parts of the University of London go their separate ways. Is that what is intended?

The peculiar characteristic of the damage that academics repeatedly inflict upon themselves is that it is always done at the eleventh hour (and sometimes the thirteenth). In London, an acceptable solution may already be beyond reach. But that possibility does not mean that it is pointless to look for a better solution. A few clarifying assumptions are obvious starting points. The Imperial College of Science and Technology, in all but name a university in its own right, is rightly acknowledged to be too important to be enmeshed in the financial administration of the University of London (whence the convention that it deals separately with the University Grants Committee); it should now be invited to go its own way. The London School of Economics, smaller but equally distinguished and important, should be given some spell of time, say ten years, to grow at the expense of the parts of the university now under threat and then also politely booted out. The rump would be three kinds of colleges — large or largeish place such as University College, smaller colleges concerned principally with the education of undergraduates, and a rag-bag of specialist institutions training veterinarians or assisting with research in outlandish languages. There is ample evidence that each kind of place is valued; what makes their administration impossible is that they are all lumped together. So why not replace them with three (or even more) different universities? When the University of Paris, as now reconstituted, consists of thirteen separate institutions, is it necessary that the largest city in Britain should be condemned to having only one?

Polish university compromise

Buffer between academics and government

Poland's new Higher Education Act, the subject of eighteen months' negotiations and confrontations — including the threat, last September, of a nationwide university strike — became law last week. The act makes several formal concessions to academics' aspirations for self-governance, but at the same time reserves to the Minister of Science, Higher Education and Technology the last word on a number of key issues.

Students, likewise, get a compromise deal — they have won the right to express their views and form associations but will not be allowed to stage protest actions since, it was explained during the parliamentary debate on the bill, "they do not belong to the community of employees".

The new act gives the minister power to monitor the resolutions adopted by university self-governance bodies and decisions taken by university rectors, and to reverse them should he decide they are inconsistent with the law. Rectors of universities and other higher colleges will be elected by the bodies themselves, the electors being either the university senate or a special collegium. The minister, however, retains the right of veto over elections — providing that he registers his objection within fourteen days.

On these and other questions, there is to be a Main Council for Science and Higher Education to hold the ring, and whose role has been negotiated between the ministry and the Social Drafting Commission — a body representing academics, students, scholars, the trade unions and the Church, which at one stage proposed its own draft bill. The council is in effect a revival under a new name of a moribund body meant to act as an intermediary between academics and the ministry.

The new act gives the council the status of the supreme elective organ of the academic community, with powers to decide on the direction of research and the organization of studies.

The Social Drafting Commission, however, had sought even wider powers for the Main Council. In the event, as Professor Zbigniew Resich, the head of the commission, said last week, "a whole series of important powers" had passed into the minister's hands.

Another major issue has been the dismissal of academic staff. The Social Drafting Commission, said Professor Resich, had considered that a lecturer's contract could be terminated only after disciplinary proceedings. In its final form,

the act now contains other possibilities, including the dismissal of junior members of staff (up to and including assistant professors) after two "negative assessment reports". It is not yet clear who would make these assessments, although it seems that the rights of the professorial body will be safeguarded, since any such dismissal must have the consent of the Council of State. But there is some evidence that "assessments" by bodies outside the university are envisaged, another limitation of "self-governance".

Similarly, although the act formally acknowledges the principles of academic freedom, including differences of "world outlook", implementation of these principles, in undergraduate teaching at least, will be constrained by what Professor Resich called "strong emphasis" in the act on the "need to teach and bring up youth in accordance with socialist principles".

These other features of the act are intended to emphasize the state character of

higher education in Poland, according to the minister, Dr Benon Miskiewicz. He himself was "following with anxiety" the participation of students and academics in illegal demonstrations, and stressed the role of the universities in ideological and political education.

His concern was timely; in Warsaw alone, six higher education establishments were represented in the arrests following last week's demonstrations. Students were equally prominent in the demonstrations in Szczecin Gliwice and in the Gdansk-Sopot-Gdynia conurbation.

These demonstrations, which took place in spite of the threat of heavy summary sentences under the emergency regulations, clearly reflect the students' bitterness at the banning of the Independent Students' Association (NZS) and the loss of many of the privileges granted in 1981. The new Higher Education Act is hardly likely to assuage their resentment.

Vera Rich

Contraction for London university

A proposal by a Joint Planning Committee that the teaching of science in the University of London should be concentrated on five sites was accepted by the senate of the university at its meeting on 29 April. No timetable for the concentration of the university is provided.

This is the university's response to the prospect of an income that will have shrunk by about 15 per cent between 1981–82 and 1983–84, and to earlier reports by the six committees set up to recommend rationalization of teaching in six subject areas. Most committees recommended that teaching should be concentrated at fewer sites.

The planning committee, however, says that these reports reflect "inadequate information" and make comments "that cannot be justified in the light of the facts". In practice, the committee responsible for biological sciences, under its independent chairman Sir James Beament, began last October by assuming that student numbers need not fall as sharply as had been decreed. The committee's second attempt under a second chairman follows the general pattern of recommending concentration on five sites.

The joint planning committee itself is hoping for some relief from the rigours prescribed by the University Grants Committee (UGC). When added together, the recommendations of the six subject committees on student numbers for the year 1983–84 imply that student numbers would then be 1,000 greater than the UGC target. The planning report says that the grants committee has been asked to agree to a larger target number or, alternatively, that the university should have an extra

year in which to make the adjustment.

On research support, the planning committee applauds the study now being carried out by the court of the university of the suggestion that financial allocations to the separate London colleges should deal separately with teaching and research. This proposal, put forward last November by Sir James Lighthill, Principal of University College, is regarded as a way of helping to avoid the consequences of the erosion of the dual-support system in the past few years, and also of making explicit the reasons why some colleges in London appear to be dealt with more generously than others. But it seems unlikely that anything will come of this before the court of the university shares out the funds available for the next academic year.

The planning committee's recommendations are that undergraduate science teaching in the university should be concentrated at five sites, where all science subjects would be taught. There are similar proposals for concentration in the arts and the humanities.

The next step in the saga of the University of London will come on 19 May, when the court of the university is to meet for a preliminary assessment of how to divide next year's grant. The exact amount will not be known until the university has heard from UGC at the end of May.

Opinion among senior members of the university is divided about the extent to which the new policy will influence next year's distribution of funds within the university. Some think that the court will be bound to distribute what money there is as the planning committee says it should. Others hold that the university will have to respect the autonomy of its colleges. ●

Paris computer centre

Orsay protests

The *Université de Paris-Sud* at Orsay, one of the principal centres of physics in France, has withdrawn from a plan to assist in the installation and development of the "largest" commercial computer in the world, a Cray-1. The computer is to be installed on the new site of the *Ecole Polytechnique* just a mile or two from the university, but Paris-Sud will have nothing to do with it (except to make a little use of it, once it is running). Orsay physicists have demonstrated over the matter and, at the same time, the scientific director of Paris-Sud's own computer laboratory, *Paris-Sud Informatique* (PSI), has threatened to resign on 20 June unless the government agrees to keep his own laboratory going.

The underlying issue is tangled, and the tangle goes right to the top. *Paris-Sud's* president, physicist Roland Omnès, is going to see the minister of education, Alain Savary, about the matter. But a remark that funding for PSI might not be renewed came from the ministry of research and technology, which is attempting to establish a computer policy for universities and research centres; and the two ministers, Savary and the research minister Jean-Pierre Chevènement, are not the best of friends.

The ministry of defence is also involved, both through its control of the *Ecole Polytechnique* and through the long-expressed wish of the military to purchase a Cray-1 (reportedly vetoed by Washington). So is the ministry of industry, which is trying to revitalize the French computer industry. Given the ever-present IBM lobby and the traditional hostility between the *grandes écoles* (such as the *Ecole Polytechnique*) and the universities, the result is a highly political problem.

At root, however, the dispute is technical. Cray-1 is a "vectorial" computer, whose great power depends on the parallel processing of a series of numbers. It is thus stretched for linear algebra (matrices and so on) only where the sizes of arrays are very large. Physicists generally do not need such arrays (except for modelling large systems such as the atmosphere) and for them the Cray-1 is far from being the most cost-effective machine. So when PSI computer engineers went to the *Ecole Polytechnique* to collaborate over Cray-1 development, they found themselves in conflict with *polytechnique* interests — which were, largely, in engineering and systems modelling. This led to personal conflict and ultimately Omnès withdrew the PSI team from the collaboration.

Since then the atmosphere has become even more heated, with PSI staff claiming, in effect, that the *Ecole Polytechnique* staff are in the hands of IBM; and indeed, the *polytechnique* has scrapped plans to link the Cray-1 to a French-built CII-Honeywell-Bull EBS-8 (a link PSI was to

Schering-Plough take \$30 million plunge

Palo Alto

DNAX Ltd, a small Californian company best known for having some pioneers in genetic engineering on its scientific advisory board, seems likely to be acquired by the New Jersey pharmaceutical company Schering-Plough for \$29.4 million. Schering-Plough already owns 13 per cent of another leading biotechnology company, Biogen SA.

Schering-Plough is to pay \$19 million in cash and is offering 340,000 shares of its common stock valued at approximately \$10.5 million. The common stock is to be distributed over 5 years under an incentive plan to the 18-member advisory board and staff of 40 scientists. Preferred shares of DNAX are at present owned by a Swiss group and a US venture capital unit of Elf Aquitaine, the French oil company.

Founded in 1980, DNAX has not so far launched any products or filed any patents. Its scientific advisory board boasts three Nobel laureates: Paul Berg and Arthur Kornberg, both of Stanford

University, and George Palade of Yale University.

Alejandro Zaffaroni, DNAX's founder, hopes that the company will eventually manufacture the protein antibodies made by the human body when it is combating infections and inflammatory diseases and allergies, and hopes to have specific macromolecules ready for testing in four to five years.

Zaffaroni is also founder and president of Alza, a company specializing in the delivery of drugs to target organs within the body. Alza's common shares in DNAX, valued at \$5.3 million, will be bought by Schering-Plough in its acquisition of DNAX.

In his search for a suitable buyer for DNAX, Zaffaroni says he approached more than 20 companies in Europe, Japan and the United States. Schering-Plough's purchase of DNAX will be the first instance of a major pharmaceuticals company acquiring an entire biotechnology firm.

Charlotte K. Beyers

work on) and will go for an IBM instead. The final crunch was a comment made at a meeting by a ministry of research and technology official, who said that Chevènement would probably not sanction the renewal of PSI equipment (a ten-year-old UNIVAC), and that Orsay physicists should use the Cray-1 "only three kilometres away".

Omnès, however, is calm. He believes official thinking is moving his university's way, and that this is not the time for a big public campaign. Nevertheless, some of his more vocal staff disagree with him. Nor does time seem to be on his side. A ministry of research spokesman said this week that it did not want to replace university computers in response to "piecemeal" pressure, and that it would take "several months" to establish a coherent policy. In the meantime the (part-time) scientific director of the Orsay computer centre may have resigned and some of his full-time technical staff, uncertain about their future, may have begun to look for other jobs. That — ultimately — would leave one of France's leading laboratories without a computer.

Robert Walgate

Law of the Sea

Soviet enterprise

In advance of the inconclusive end to the UN conference on the law of the sea, the Soviet Union last month passed enabling legislation to cover the granting of prospecting and exploitation licences to Soviet enterprises wishing to engage in undersea mining beyond the continental shelf.

The decree of the Supreme Soviet now

published describes the measures as "temporary" and stresses that the Soviet Union "is and will continue to be in favour of the settlement of urgent problems of the legal regime of the world ocean on an international basis . . . taking into account the legitimate interests of all states". In the absence of any such legislation, however, the Soviet Union has been "compelled" to protect its interests by granting off-shelf licences to Soviet mineralogical concerns.

The decree was clearly an answer to Western proposals on off-shelf mining. In particular, the head of the Soviet delegation to the law of the sea conference, deputy foreign minister S. P. Kozyrev, categorically rejected Western proposals to grant licences on a "first-come first served" basis. This, he said, was a principle "widely employed during the colonial enslavement of countries and peoples".

Commenting on the new decree, the daily *Socialisticheskaya Industriya* stressed the interest of US companies in off-shelf strategic minerals which, it alleged, had pressurized President Reagan to renege on earlier understandings. The new decree has therefore been interpreted by some commentators as simply a tactical move to strengthen the Soviet position at the bargaining table.

The provisions of the decree, however, are considerably more detailed than would be necessary for such a purpose. It goes into considerable detail about safety zones around deep drilling installations, the notification of temporary navigation hazards, disputes over rights with the "agencies of foreign states" and the setting up of a special fund under the Council of Ministers to which Soviet licence-holders must pay part of their profits.

Moreover, a time-lag has been built into the decree, presumably to allow the Soviet Union to acquire the necessary technology. No off-shelf operations, it states, can begin before 1 January 1988.

Several clauses deal with possible non-Soviet participation in off-shelf operations. This would presumably be of two types — with developed and with developing countries. One likely partner in the first category is East Germany, with which the Soviet Union (together with Poland) is already linked in the Baltic drilling consortium *Petrobaltik* to which the Germans have made a considerable

technological contribution. The decree, however, stresses the Soviet Union's role in "assisting" its future partner in the development of technology and the training of personnel, suggesting participation by developing countries. (This would fit in with Kozyrev's rejection of the "colonial" overtones of the Western proposals.)

The developing country would in the first instance presumably provide a shore base and other facilities. Any minerals recovered, according to the decree, would become Soviet property, though in some cases, some or all of them could revert to

the partner state. The concept of the Soviet Union building up Third-World off-shelf technology would, however, become particularly relevant if an international agreement were drawn up dividing the off-shelf zone between the interested countries. In that case, the "temporary measures" now announced would automatically lapse. But existing collaboration agreements between the Soviet Union and developing countries on mining activities normally provide for the Soviet Union's initial investment in the technology being repaid over several decades in appropriate minerals. **Vera Rich**

Conference all at sea

Washington

The largest ever diplomatic conference of the United Nations, with 160 governments represented, ended on 30 April in New York with a huge majority for a draft treaty on the Law of the Sea. But the participants in this eight-year-long conference have not gone home in triumph. Instead, they hope to limit the damage that has been done.

What went wrong? In the end, the United States voted against the treaty, as did Israel, Turkey and Venezuela. Britain, West Germany, the Netherlands, Belgium, Spain and the entire Eastern bloc abstained. Ironically, even the dissenters think the draft is a splendid basis for a treaty. It resolves several complex issues, such as navigation through narrow straits.

The stumbling block turned out to be the issue well advertised in the past few years: the rights of private companies to minerals from the deep sea bed. Both the United States and the Soviet Union were affronted. The new Administration in Washington had spent a year trying to decide how to amend the clauses on international access to oceanic minerals that the Carter Administration had negotiated. The Soviet Union, by contrast, objected to a form of words seeming to cast doubt on its freedom from free enterprise.

What happens now is anybody's guess. The procedure requires that governments wishing to sign the treaty must do so by December. Thereafter, the treaty will become international law when a specified quota of signatories has formally ratified it. The flaw is that because the draft of the treaty was not approved without dissent, it is open to any member of the conference to insist that the text cannot be a basis for international law because it was disputed.

So who shoulders the blame? The outcome of the conference was a failure for the diplomats. The bizarre handling of the United States negotiating team may have been crucial.

The Law of the Sea session just ended lasted just eight weeks, and brought to an end the year-long spell during which the United States Administration had put the negotiations on ice.

Private seabed mining has been the only stumbling block. Ideologically, the Administration objected to those parts of the text that would give developing countries a say over the private development of seabed minerals. But among the six objections listed by the United States when it returned to the negotiations in January were the provision that the treaty could be amended by three-quarters of its signatories without the US Senate's agreement, and the inclusion of the Palestine Liberation Organization among the members of the conference.

After the event, US negotiators claim that they were not given enough time to negotiate the real issues. The conference chairman, Tommy Koh of Singapore, made them stick to a prearranged timetable. The deadline of 29 March had early on been fixed for the filing of committee reports on which a final draft would be drawn up. By then, however, the mining committee had not agreed a resolution on seabed mining, while the constitutional objection to changing the treaty without consent of the US Congress had not been discussed.

Part of the difficulty seems to rest with the makeup of the US negotiating team. Its leader was James L. Malone, an assistant secretary at the State Department whom the Administration was stripping of his responsibility for nuclear non-proliferation. Malone's chief aide was Leigh Ratiner, a former government seabed mining expert working for the mining industry.

Ratiner was a free-wheeling negotiator and a veteran of earlier sessions at the law of the sea conference. Malone was by comparison a novice. Observers say that Malone would often say one thing and Ratiner another. The result would be either brilliant or confusing, depending on the observer. The delegation was under constant pressure from the mining lobby, which insisted that any sell-out to the conference would be denounced when the treaty came up for ratification in Congress.

The United States' demands for revisions of the treaty, circulated before the beginning of the last session in March,

were quickly followed up by detailed textual proposals — and as quickly rejected as too sweeping. A compromise proposal, put forward by 11 states including Australia, Canada, Denmark, Finland, Iceland, Ireland, New Zealand, Norway, Sweden and Switzerland, was quickly rejected by the United States as inadequate but later recognized to be a basis for discussion. By then, however, it was too late.

What will happen now is far from clear. The ceremony at which the treaty will be opened for signature has been arranged for Caracas in December, but the host country, Venezuela, voted against the draft last month because it fails to resolve the dispute with neighbouring Colombia.

No state is bound by the vote it cast on 30 April. Abstaining states or no-voting states could still attend the signing ceremony. But Israel's no is unlikely to change, as the PLO no seems likely to stay in the final text. The Reagan Administration could not change the US vote without an ideological turnabout, which seems unlikely.

Ironically, the Administration's adamant defence of rights to ocean mining may have brought about a situation in which no mining is possible. Tommy Koh has threatened to sue any country that proceeds unilaterally to mine outside the framework of the Law of the Sea. So, if the United States, and other mining states such as the United Kingdom, remain out of the fold, they may see their mining ventures flee to other countries such as Japan.

Meanwhile Elliot L. Richardson, Law of the Sea negotiator for the Carter Administration, believes that the outcome of the conference is "sad". Richardson said that "ideologues" in and outside the Reagan Administration prevented the US negotiating team from being flexible.

The Administration, he said, "wasted" a year deciding whether to continue negotiating, and so spent too little time preparing for the talks themselves. At the meeting at which the US team agreed to produce specific draft articles, it had run out of time and had to include many items that it knew were unacceptable. Similarly, Richardson believes the US team should not have responded so negatively to efforts by a group of eleven countries to negotiate compromises. **Deborah Shapley**

UK nuclear power

Bye British

At last the British nuclear power show is on the road. Advanced gas cooling is an also-ran in the UK Central Electricity Generating Board (CEGB)'s "statement of case" for the construction of a pressurized water reactor at Sizewell in Suffolk, published with due ceremony yesterday (Wednesday).

The price of the British 1,110 megawatt-electric power station is now estimated to be £1,147 million, excluding the initial charge of fuel, compared with £1,590 million for an AGR. The ratio is thus 72 per cent, substantially but not embarrassingly more than the 60 per cent Dr Walter Marshall, chairman of the PWR task force, was hoping for.

The task force was set up to control the

cost of safety systems on the British PWR, which a year ago were escalating because of CEGB requirements to keep radiation exposure levels for PWR workers down to those expected of the cleaner gas-cooled reactor. Then there were fears that a British PWR might cost as much as an AGR. In the event, CEGB has maintained its safety requirements, adding elements — such as a larger containment building and a double skin — to the task force design; but the task force effort has shown up in the final price.

An equivalent coal-fired station would cost about the same as the PWR (£1,080 million), says CEGB, but fuel costs would be so much higher with coal that a PWR would give a net saving of some £500–£1,100 million over its 35-year life.

But these few lines give no indication of the scale and density of CEGB's argument which in the statement of case, the accompanying pre-construction safety report, the reference design and the supporting documents amounts to 2½ hundredweight (125 kg) of reading matter.

PWR objectors will now have eight months to study the case, before the public planning inquiry which is to be held at the Maltings, Snape, in Suffolk, in January 1983 — and they will need all of that time. Moreover, CEGB promises a supplement in two months' time which will give a sensitivity analysis (or estimate errors) on the economic elements of the report, and in late June or early July the Nuclear Installations Inspectorate is due to publish its reaction to CEGB's design. The opposition could thus find itself flattened by mere weight of words.

The report, and preparation for the inquiry (including fees to counsel), will cost CEGB some £5–10 million, it is estimated. Some copies will be made available to *bona fide* enquirers, 15 will be placed on public display and others will be sold at £250 a time.

The next crucial task for CEGB is to define the board's procurement policy — where it will buy the components and design work for the reactor. The board estimates that 90 per cent of the PWR cost will pay for work done in Britain — but since the nuclear component (the nuclear steam supply system) amounts to only 10 per cent, CEGB is clearly not ruling out the possibility that the whole of the system might be bought abroad.

The reactor pressure vessel will certainly be bought abroad. Tenders are now out with Combustion Engineering of the United States and Framatome of France. For the rest of the system (such as the steam generators), discussions are beginning with potential British suppliers. Brian George, CEGB's PWR director, said it would be "too strong" to say that the board saw no reason why it should not buy British. In other words, there are good reasons — of price and confidence — why the board should buy foreign, and it will be up to British nuclear industry to convince CEGB otherwise.

Robert Walgate

Alternative energy

Winding up

Waalre, The Netherlands

The Dutch are again turning to wind as a source of energy. At the end of the last century, more than 9,000 windmills were pumping water or grinding flour in the Netherlands, and if the present emphasis on wind energy continues, by the year 2000 there will be more than 2,000 megawatts of electricity produced from wind power. This is equivalent to about 13 per cent of the country's present requirement for electricity, which is not expected to increase much in the next few decades.



"Looks like the wind's dropped again"

Approximately two-thirds of the wind-derived electricity at the end of this century is expected to come from purpose-built wind-energy parks, with the rest from up to 15,000 smaller individual turbines. There are already 50 small turbines in use, manufactured by 7 Dutch and 14 foreign companies. Until now the growth of the use of wind energy has depended on the enthusiasm of local authorities, but the end of the national five-year research programme and the move into a development programme should see a further boost to wind energy.

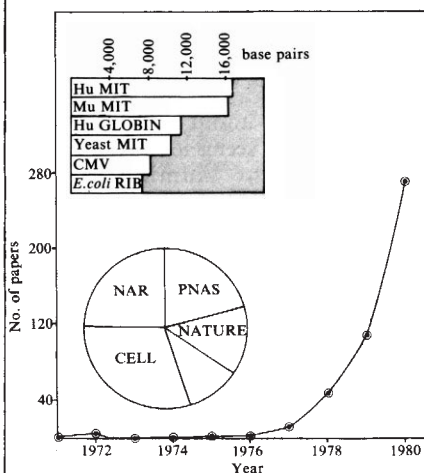
The national research programme saw the construction of a 300-kilowatt test turbine, and the new programme will initially involve the building of a wind-energy "farm" with forty 250-kilowatt turbines as a prototype for much larger central electricity generating areas. The cost of this first "farm" is estimated at 40 million guilders (£8 million), half to be paid by the government and half by the public utilities companies.

The government is encouraging the construction of wind turbines by providing a subsidy of up to 40 per cent, which can be paid back over 10 years. The national development programme will also involve the design of a large 60–70 metre commercial wind turbine. Further research is to be done on tip-vanes and a multiple wind turbine is under development, comprising a tower with three cross-bars, each of which will support two rotors, giving a total capacity of 320 kilowatts.

Casper Schuurin

Sequences add up

The first newsletter of the newly established nucleotide sequence data library of the European Molecular Biology Laboratory (see *Nature* 15 April, p. 596) lists 565 entries and is claimed to be fairly complete up until 1980 and perhaps half complete for 1981. Four journals, *Cell*, *Nucleic Acids Research*, *Proceedings of the National Academy of Sciences* and *Nature*, dominate the entries. The data base contains nearly



600,000 nucleotides, with 18 continuous sequences of greater than 5,000 nucleotides. The longest of these, at 16,569 nucleotides, is that of the human mitochondrion (HuMIT). Mouse and yeast mitochondria and a human globin gene cluster also provide sequences of greater than 10,000 nucleotides and cauliflower mosaic virus (CMV) and *Escherichia coli* ribosomal genes are not far behind. The mouse provides the biggest number of entries, followed by man. In both cases the sequences are largely of immunoglobulin, globin and interferon genes. The newsletter is available from G. Hamm, EMBL, Postfach 10 22 09, D-6900, Heidelberg, FRG.

West German waterways

Canal in limbo

Heidelberg

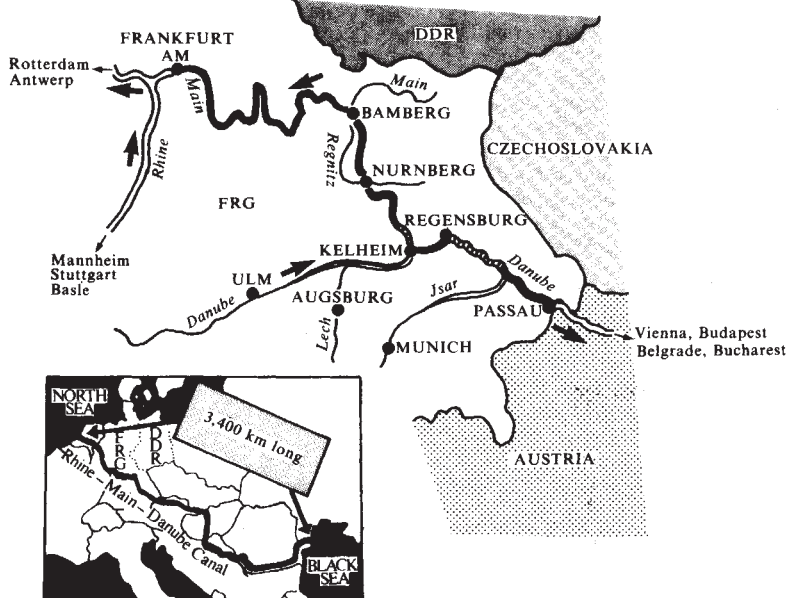
Of the 3,500 km of waterway linking the North Sea to the Black Sea, only 55 km of the Rhine–Main–Danube Canal and 59 km of canalization on the Danube are still incomplete. However, as the work forges on, money is tighter, and controversy about the impact of the canal increases.

The first attempt to conquer the European watershed by joining the Rhine to the Danube dates back to Charlemagne. But it was not until 1855 that, under Bavarian King Ludwig I, a 15-m wide canal with 100 locks was opened. This never competed successfully with the railways and the last 120-ton barge passed through in 1945. Contracts from 1921 and from the 1960s commit the federal government to two-thirds and Bavaria to one-third of the costs of the present canal from Bamberg on the

million per year for running the canal might benefit the area more if spent differently. More jobs could be lost in farming and tourism than are created by the canal.

The plans also involve pumping water at 15 m³ per second up from the Danube to flush out the severely polluted Regnitz and Main and provide cooling water for the four new power plants planned for the Main valley. Without the canal, a DM 500 million pipeline would be needed. But, integrated energy planning and improved waste treatment could meet these needs.

To the 55-m width of the canal will be added a huge parallel service area which in places will carpet the valleys of the Altmühl and Sulz from wall to wall. After canalization the Danube will deepen its bed, lower the water table and drain surrounding wetlands. These vast landscape changes are the target of mounting environmentalist attack. The Altmühl valley and the wetlands north of the Sulz and east of Regensburg are not only exceptionally



Main to Kelheim on the Danube.

Even by 1976, estimates showed that on each deutschemark invested, the profit would be only 50 pfennig. Total German water freight is in any case falling off: estimates for the canal were 14 million tonnes per year in 1969 and now stand at 3 million tonnes per year. Most of this would otherwise go by the *Bundesbahn*, which is predicted to lose DM60–120 million per year. Worse, German national carriers could face competition from state-subsidized East Europeans not only on the new waterway, but also in the crucial Rhine system. Furthermore, north German seaports could suffer when the canal links Austria, Hungary, Bulgaria and Romania directly to the Rhine delta.

Bavaria considers the project essential to combat local unemployment and underdevelopment and that the DM 3,150 million already invested should not be jeopardized. However, the DM 1,750 million for completing construction and the DM 12

beautiful, but also contain internationally significant biotopes. The destruction of the migrant wintering and resting grounds, the habitats of 46 endangered bird species and 13 endangered fish species, is contrary to the spirit of both Bavarian and federal laws framed in the 1970s to protect fast dwindling wildlife resources.

These serious environmental objections apart, the hard core of the current controversy is economic. In September 1981, while West German Chancellor Helmut Schmidt reassured Kreisky that work on the canal would continue, federal transport minister Hauff, on the strength of his latest reviews, was talking about "the dumbest project since the Tower of Babel". In January the cabinet instructed Hauff to ask the Bavarian Administration for talks about halting or reducing the scheme. Bavaria is holding firmly to its contracts, but Hauff and Strauss will meet this month. Meanwhile, there is nothing in the federal kitty for next year. **Sarah Tooz**

Kew Gardens

Changes ahead

In the cold climate of recession, the Royal Botanic Gardens at Kew in Surrey continue to flourish. Today (13 May) the Queen will re-open the Temperate House after a 5-year renovation programme costing £1.6 million. Building of this vast glasshouse, designed by Decimus Burton for the 3,000 different types of plant from the warm, temperate regions of the world, began in 1860.

The Royal Botanical Gardens may themselves soon undergo a historic restructuring. Since 1903, the institution has been part of the UK Ministry of Agriculture, Fisheries and Food (MAFF) but now talks are being held between officials from the ministry and Kew Gardens to transfer management to a body of trustees. Organizations with an interest in the gardens have been asked to submit their views by 21 May.

The change to trustee status would be largely managerial — the gardens would continue to be sponsored by MAFF. But this arrangement would permit a degree of administrative autonomy the lack of which has irritated past directors of Kew. The new board of trustees, many of whom would be scientists, would take decisions on the direction and development of policy. In April, Mr Peter Walker, Minister of Agriculture, Fisheries and Food, told the House of Commons that the present arrangements "are not ideally suited to the management of an institution which combines the functions of curation, research, advice and instruction and public amenity". At present the director, Professor Arthur Bell, is advised by a 12-strong Scientific Advisory Committee drawn mainly from university botany departments. Members of the Forestry Commission and the National Trust sit on a consultative panel at Wakehurst Place, the satellite garden in Sussex.

Kew Gardens have not experienced cuts as such but staff has been reduced through natural wastage from 480 in 1979 to the present level of 440 — a rate which compares with cutbacks at the ministry itself. Total running costs are £6 million, of which £3 million is accounted for by salaries.

Like London Zoo, Kew is a blend of scientific institution and tourist attraction. With an entrance fee of only 10 pence, and over 1 million visitors a year, however, the gardens could not survive without exchequer support. Some 3,100 genera of 340 families are represented in the living collections at Kew, and there are 5 million dried specimens in the herbarium. The physiology section at Wakehurst Place now incorporates a seed bank. But the aesthetic way of displaying plants of scientific and economic interest may obscure the extensive scientific work that the institution performs. **Jane Wynn**

CORRESPONDENCE

The Johns Hopkins and India

SIR — In the news section of your 11 March issue, K. S. Jayaraman from New Delhi makes several serious and erroneous allegations regarding the Johns Hopkins University and its participation in research projects in India. I believe it to be important that the record be corrected.

For many years, scientists from the Johns Hopkins University collaborated with Indian colleges in two major projects. One was the pioneering research conducted from 1961 to 1974 at the Narangwal Rural Health Research Centre in the Punjab. The purpose was to design and evaluate practical approaches for the delivery of primary health care. A series of projects under the Indian Council of Medical Research involved a large group of Indian scientists and a few foreign scientists under the direction of Professor Carl Taylor. These studies served as a prototype for other similar health services research subsequently conducted in many countries throughout the world. Professor Taylor, a member of the Institute of Medicine of the National Academy of Sciences and a frequent consultant to the World Health Organization, the World Bank and many other national and international organizations, is universally recognized as an authority in the complex problems of primary health care delivery. He was born in India and has spent half of his life living and working in the villages of the subcontinent. Contrary to Mr Jayaraman's report, the Indian government has officially confirmed that he has never been expelled from India nor has he been forbidden to come to India or any other country in a professional capacity. His visits to India since the Narangwal Project closed have been in response to official invitations. The allegation that the purpose of the Narangwal study was to provide a base for spying on the Halwara airbase is wholly false, indeed preposterous.

The second project, based at several national institutes in Calcutta, was under the direction of the late Professor Frederik Bang. It was one of a group of International Centers for Medical Research and Training which were funded by the National Institutes of Health. With Indian scientific colleagues, a large number of important investigations were conducted dealing with infectious diseases and nutrition. Indeed, many of the fundamental studies which have led to the now universally used methods for treatment of cholera and other severe diarrhoeas with oral fluid therapy were made by the Calcutta group. This laboratory and its scientific staff had no connection whatsoever with the Fort Detrick biological warfare laboratory. There was also no link with the US Navy except for informal, collegial scientific exchanges with the Naval Medical Research Unit, then based in Taipei, which was similarly actively engaged in endeavouring to find more effective and practical methods for the treatment of cholera.

Since its founding in 1916, the Johns Hopkins School of Hygiene and Public Health has been deeply involved in education and research in international health problems in Baltimore and in many countries throughout

the world. It continues today as one of the largest institutions so concerned with international health. Its numerous alumni hold eminent positions in national and international organizations around the world. The false aspersions cast on the institution by Mr Jayaraman are deeply resented.

D. A. HENDERSON
(Dean)

*The Johns Hopkins University,
School of Hygiene and Public Health,
Baltimore, Maryland, USA*

Ban the billion

SIR — A. J. Southward (*Nature* 294, 215; 1981) writes: "... estimated six billion tons of oil that reaches the sea each year", as he comes from Plymouth, Devon, we have a right to expect that "billion" means 10^{12} ; but in context that is surely impossible. We are therefore left to wonder whether he writes in American from an English address. Unfortunately, even " 6×10^9 tons" of oil per year flowing into the sea sounds fantastic, given that the world's production of crude oil is about 3×10^9 tons per year. Does that mean that most of the amount talked about is natural seepage? The rest of the article makes that sound implausible. We are left with the uncomfortable feeling that, as the author has introduced one uncertainty of a factor of 1,000, he may just have got the figure wrong.

The confusion is worse confounded by the difficulty of distinguishing spoken "m" and "b", especially on the telephone. As a result, "million" and "billion" are frequently confused, particularly in newspapers. The uncertainty is thus increased to 10^6 .

Sir, could I prevail upon you simply to ban the word "billion" from *Nature*? It seems to me unnecessary; given the enormous uncertainty, it would be much better simply replaced by " 10^9 " or "thousand million".

EDWARD EISNER

*Department of Applied Physics,
University of Strathclyde,
Glasgow, UK*

A. J. SOUTHWARD REPLIES — *Mea culpa*: Professor Eisner is quite entitled to his little bit of fun. The word should have been million — 6 million tons of oil per annum into the sea. I agree that the word billion needs replacing.

PhD applications

SIR — The leading article in *Nature* of 15 April (p.592) implies that "promising people from less favoured universities" find it hard to enter PhD courses because the quota system of awards puts graduate student selection into the hands of the luckier university departments.

I expect that supervisors usually want to choose the most productive and biddable graduate students quite independently of their college of origin. If this is so, the main barrier to students moving to a new institution for their graduate studies must be inadequacy of

information about projects and places available elsewhere.

For several years this department (which has always had a strict policy of not allowing its own honours students to continue as PhD students here), has organized a scheme for exchanging this information amongst about 80 UK life science departments — mainly in biochemistry, molecular biology, microbiology and cell biology. The scheme also promotes a common application form and loosely coordinated timetable for selection.

This year, for the first time, we have published the *Compendium of Research Topic Outlines*, which forms part of the scheme, as a single volume of over 400 pages, with name and subject indexes. The first indications are that this has proved very useful to students in finding out where they can carry out the kind of research of most interest to them. In the future, the availability of this compendium to libraries and career advisers should make the system of greater use to students in non-participating departments.

I shall be glad to answer enquiries from anyone who thinks that the compendium, or the scheme as a whole, could be useful to their students.

I think that many of the participating departments now find the scheme indispensable, as we do. For this reason I have never understood why physical chemical and social science departments do not cooperate in a similar way.

A. F. W. COULSON

*Department of Molecular Biology,
University of Edinburgh,
Edinburgh, UK*

Man mismeasured

SIR — Scientifically I am certainly no supporter of the main thesis of Stephen Jay Gould concerning punctuated equilibrium. Thus I might be expected to be in favour of (or at least to watch passively) virtually anything that helps to reduce Gould's immense influence and effectiveness, especially if prepared by an outsider to the evolutionary arguments. But in truth, I found the personal attack on Gould by your reviewer of *The Mismeasure of Man* (*Nature* 8 April, p.506) positively nasty, unprofessional, with but slight reference to substantive issues (and then often granting Gould's correctness), and just plain ugly. Who is Mr Blinkhorn to tell us that the book has "the routine flavour of Radio Moscow news broadcasts", as though in and of itself even if it were true (which is not in any way documented), it should damn the book for ever. My reaction is that if Blinkhorn is correct, more of us should join him in listening to those radio waves. Blinkhorn ends by admiring Gould's skill in presentation, and adding "but what a waste of talent". For his part, Blinkhorn has shown us a remarkable lack of skill, unfortunately presumably making full use of his talent.

THOMAS J.M. SCHOPF

*Department of Geophysical Sciences,
University of Chicago,
Chicago, Illinois, USA*

NEWS AND VIEWS

Evolution of bowerbirds' bowers: animal origins of the aesthetic sense

from Jared M. Diamond

In September 1872 the explorer Odoardo Beccari became the first European to ascend the Vogelkop mountains of New Guinea and to observe there the Arfak tribesmen in their villages. Among the many sights that impressed Beccari were the beautifully decorated little huts scattered in the forest. Each hut was in the form of a cone up to eight feet long and four and a half feet high, so tightly woven of sticks as to be rain-proof. In front of each hut was a little garden, consisting of a mossy lawn decorated with over a hundred red, blue, brown, black, yellow and green objects carefully grouped by colour. The objects that Beccari saw were natural and collected from the forest: fruits, flowers, mushrooms and beetle skeletons. The displayed flowers were fresh and were removed and replaced as they faded. A century later, when Europeans had become regular visitors to the Vogelkop, the decorations included coloured man-made objects, such as shotgun cartridges and yellow cardboard Kodak film cartons.

Was this a quaint native custom of the sort that draws anthropologists to New Guinea? Not at all. What caused a sensation when Beccari^{1,2} published his account of these huts was that they were constructed, decorated and used by a bird, the Vogelkop Gardener Bowerbird.

The eighteen species of bowerbirds (family Ptilonorhynchidae) of which fourteen others besides the Vogelkop Gardener build bowers, are confined to New Guinea and Australia³⁻⁵. Bower structures vary with the species and include, besides the hut of the Vogelkop species, stick towers up to eight feet high, walled avenues, fern mats and moss platforms with parapets. Almost all bowers are decorated with coloured objects such as fruits, flowers, shells and bones, the colour preference varying with the species and with the individual bird. More diverse decorative objects appear in bowers near human habitation, and have included car keys, bottle-tops, tooth brushes, clothespins, false teeth and spectacles. At

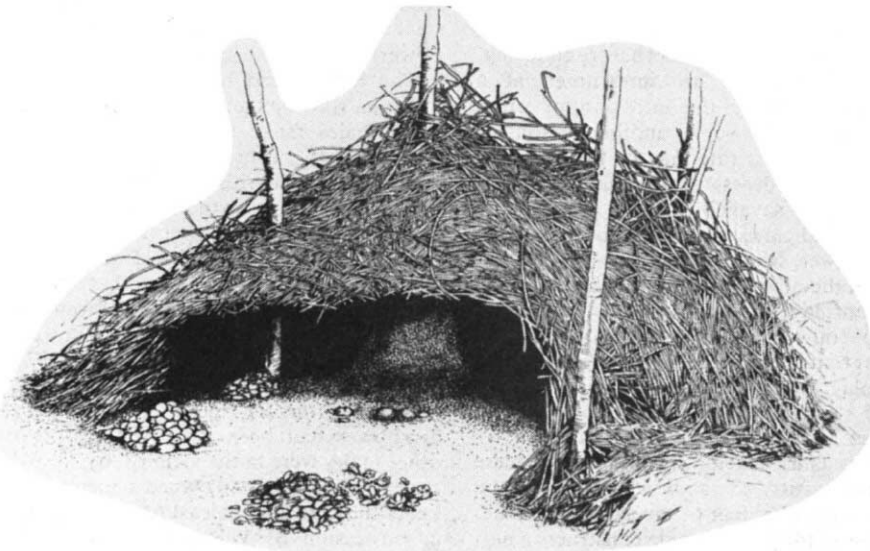
least six species use a tool to paint the bower yellow, green, blue, brown, or black with crushed fruit or charcoal, and at least four species orient the bower in a constant compass direction. Bowers are built by adult males and used as the site for wooing and inseminating females, which then fly off and do all the work of building a nest and rearing the young.

Understanding of the function and evolution of the bowers has come slowly, because the areas of New Guinea and Australia where many of the species live are remote and difficult of access. Indeed, the home grounds of a New Guinea species (*Amblyornis flavifrons*), known only from three skins sold to Lord Rothschild by a plume merchant in 1895, remained undiscovered until 1981 (see *Science* 23 April 1982), and the bowers of two other New Guinea populations (*Sericulus bakeri* and *Archboldia p. papuensis*) are still unknown. Two recent papers by Reta Vellenga^{6,7} of Castle Hill, Australia, in conjunction with an earlier paper⁸, provide the most detailed life history information yet available for a bowerbird.

Vellenga studied the Satin Bowerbird (*Ptilonorhynchus violaceus*) of south-east

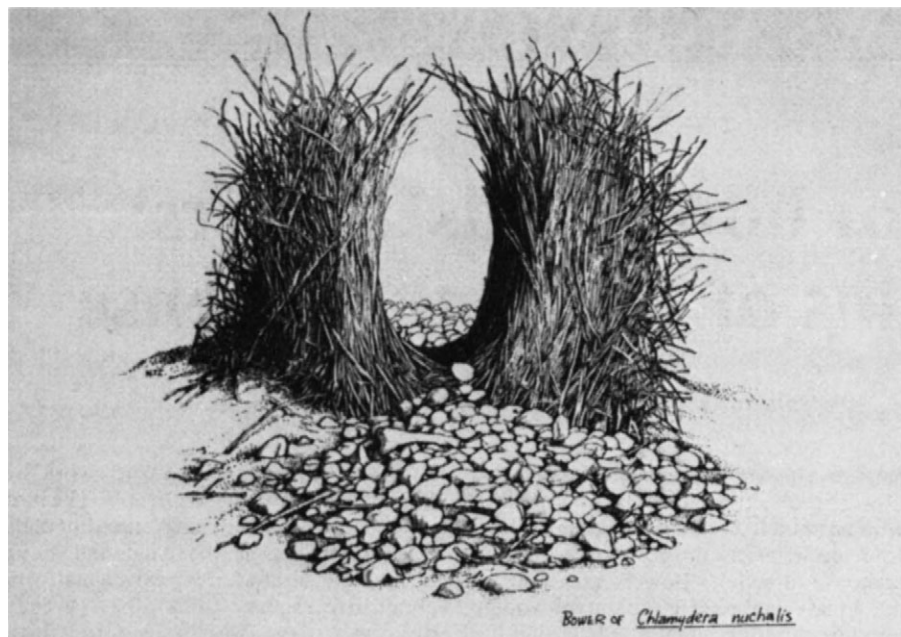
Australia, following on earlier work by Marshall³, Chaffer⁹ and Green¹⁰. The bird is about twelve inches long, the adult male glossy bluish-purple, the female and young male green. The bower is a woven platform about 10 feet square, supporting an avenue of woven sticks, with walls a foot high nearly joining in an arch over a floor, and decorated with blue or green natural or man-made objects. Perishable decorations such as flowers are replaced daily. Some bowers are left unpainted, but others are painted daily either blue, black, or green by the male, using a wad of bark as a paint brush and using crushed fruit, charcoal, or (nowadays near civilization) stolen blue laundry powder as paint. The long axis of the bower is generally within 30 degrees of north-south, possibly so that the male and female can face each other during early morning displays without either having to stare into the rising sun. When Marshall picked up a bower and reoriented it, its architect promptly demolished it and rebuilt it in the correct orientation. Between 1965 and 1971 Vellenga banded 940 bowerbirds in her backyard and followed the individual lives of many of them.

Extremely elaborate maypole bower built by a completely unornamented male bowerbird, *Amblyornis inornatus* (mountains of New Guinea's Vogelkop Peninsula). The male lacks a crest. The bower is a hut of interwoven sticks built about one or more saplings, with a lawn in front decorated by over 100 objects of up to seven different colours.



Drawings by W.T. Cooper (ref. 5).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024.



Example of an avenue bower, built by the bowerbird *Chlamydera nuchalis* (Australia). An avenue is constructed of a floor and two painted walls of sticks and decorated with many stones and bones. Mating takes place in the avenue.

According to Darwin's theory of sexual selection, intraspecific competition for mates may select for ornaments or behaviours unique to one sex through either of two mechanisms: the value of the ornaments in attracting the opposite sex or in dominating rivals of the same sex. Bowers have traditionally been interpreted in the former sense, as love parlours where males seduce females. Vellenga's observations provide ample evidence for this function. The bower-owning male Satin Bowerbird continually tries to entice females into his bower by picking up in his bill an object such as a flower or snail shell, and posturing, dancing and displaying to the female. Rarely, the female is persuaded to enter the bower and copulation ensues. Far more frequently, the female flees without mating.

Why do courting males have such a low success rate? Perhaps females scrutinize various males and bowers before choosing. Many courtships are interrupted at a crucial stage by the intrusion of other individuals. However, another reason for the low success rate may be the invariant sequel to successful mating: the male bowerbird savagely attacks the female, pecks and claws her, and chases her from the bower. Mating itself is so violent that often the bower is partly wrecked, and the exhausted female can scarcely crawl away. The courtship display can appear little different from the male's aggressive display. When a courted female is won over and starts to solicit copulation, the male often changes his mind and chases her away. Thus, a female may have to make many visits to a bower before she overcomes her fear of the aggressive male. After mating, the female constructs a nest

at least 200 yards from the bower and bears sole responsibility for feeding the young. Once the young leave the nest, the young of several females are tended by their mothers in a common nursery area. After the first year, bowerbirds outside the breeding season form separate flocks according to age and sex (females, young males and old males separate). Adult males that leave for the winter apparently dismantle their bowers first.

While Vellenga's studies confirm the traditional view of bowers as love parlours, they provide strong support for a simultaneous role of bowers in male-male interactions. Specifically, bowers, like flags of possession, may serve as symbols of males' property rights in their wars with other males. An adult male spends much time repairing his own bower, protecting it from raids by rivals, and attempting to steal ornaments or destroy rivals' bowers.

The battles and territorial shifts that Vellenga recorded among her 426 banded adult males make the European Thirty Years' War seem straightforward by comparison. When about 5 years old, male no. 85 won ownership of Vellenga's garden and maintained his bower there against intruders for 15 years. On 15 November 1967 he became ill, and other males demolished his bower, but by 28 November he had recovered and rebuilt the bower. Finally, on the morning of 26 October 1970 he became ill again and disappeared around 10 a.m. By 1:50 p.m. on that day his bower had been destroyed and two other males were in the vicinity. By 3 p.m. male no. 345, who had owned a bower 1.25 km to the west since at least April 1970, was in possession of Vellenga's garden and

began constructing a new bower around 4 p.m. Male no. 345 retained Vellenga's garden through constant fighting until defeated on 3 January 1971 by male no. 460, who had until then owned a bower 7 km distant. Male no. 345 thereupon returned to his old bower, expelled a male who had taken it over and was still there when visited three years later, while male 460 still owns Vellenga's garden at latest report. In the meantime male no. 553 took over the bower of male no. 320, 0.75 km south-west of male 345, while male no. 429 seized a bower 1 km to the south-east, and male no. 124. . .

What is the significance of all these details of bowerbird military history? As with other bird species, male bowerbirds vary in dominance, and the south-east Australian landscape varies in production of food resources consumed by bowerbirds. Superior locations such as Vellenga's garden are known to have supported active bowers for at least 30 years, though with a sequence of owners. Young males search for good sites to build a bower, and males that already own bowers spend some time searching for better sites. Thus, the old male 85 had long been dominant in Vellenga's neighborhood, and on his disappearance local dominance did not pass to a male at some different site but continued with the new owner of male 85's estate.

Dominant males directly prevent other males from wooing females by the destruction of their bowers. Young males continually try to erect rudimentary bowers in the territory of an adult male but the latter patrols his property several times a day and wrecks them. When Marshall placed 100 pieces of numbered blue glass into these rudimentary bowers one night, he found 76 of the pieces transferred to the bowers of dominant adult males by noon of the next day. Similarly, Vellenga observed a blue celluloid band to be transferred between bowers several times a day, until one male firmly wove it into his bower.

Dominant males may perhaps even affect their rival's chances through hormonal effects related to dominance and subordination. The successful male bowerbird needs not only a bower but also blue plumage. Young male bowerbirds have green plumage resembling adult females and only begin to acquire the blue plumage of the adult male in their fifth year. In many animals the acquisition of adult sex-specific colour pattern is controlled by sex hormones, and domination of a male by rivals inhibits production of male sex hormones — it is possible that the acquisition of blue adult plumage by male Satin Bowerbirds is retarded in individuals that are continually harassed and see their bower-building attempts frustrated by dominant males.

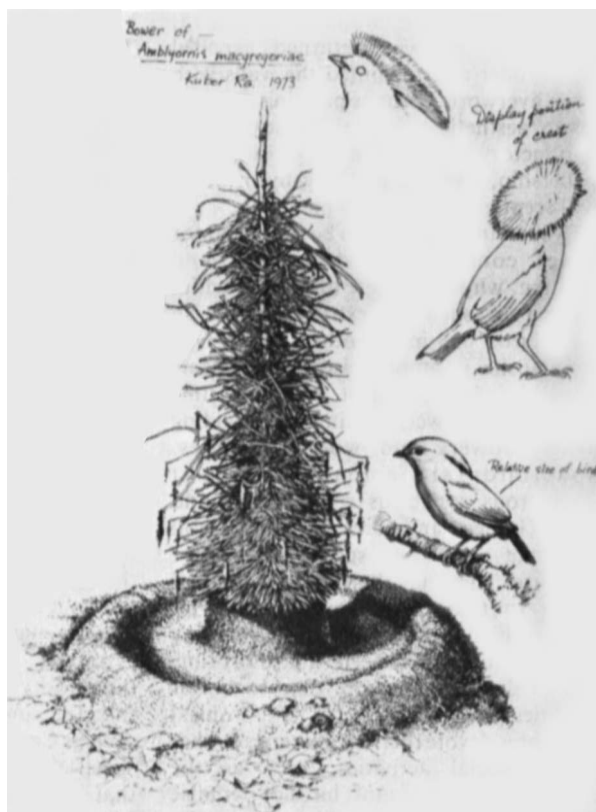
How has the evolution of bowers taken place? Gilliard^{4,13} argued that the female's attention during the male's courtship display has become 'transferred' from

male ornamental plumage to the bower and its decorations. The transfer is an advantage for the male bowerbird because it enables him to dispense with bright plumage that may attract predators. One line of evidence for this interpretation comes from comparison of courtship displays in the four species of the avenue-builder genus *Chlamydera*. In two species, *C. nuchalis* and *C. maculata*, the male has a lilac crest, which he erects and displays to the female by facing her while twisting his head to the side. In the crestless *C. cerviniventris*, the displaying male still twists his head as if to show the female a crest that his ancestors lost! But the male *C. cerviniventris* compensates by holding a spring of green berries in his bill during the display.

A second line of evidence for the 'transfer theory' is the inverse relationship, among bowerbird species, between elaborateness of male plumage and elaborateness of the bower and its ornaments. For example, among the five species of maypole builders, the male of *Amblyornis macgregoriae* has an erectile orange crest three inches long, builds a simple maypole four feet high and scarcely decorates it; male *A. flavifrons* has a gold crest of the same length, builds a similar maypole and decorates it only with three small piles of fruit of three different colours; male *A. subalaris* has an orange crest only 1.5 inches long but builds a dome over his maypole and decorates it with flowers, fruit, beetles and leaves of at least five different colours; male *Prionodura newtoniana* has a very short yellow crest but builds and decorates two maypoles up to eight feet high and joined by a bridge; and male *Amblyornis inornatus* is entirely crestless like the female but builds the elaborate bower described in the first paragraph of this article. A similar correlation exists among the eight species of avenue builders: from *Sericulus aureus* and *S. chrysocephalus*, of which adult males have golden and black plumage, decorate a two-walled bower only sparsely, and perhaps dispense entirely with a bower on occasion; to *Ptilonorhynchus violaceus* studied by Vellenga, with uniform blue male plumage and a painted two-walled bower with many ornaments; and culminating in *Chlamydera lauterbachii*, in which male and female are similarly dull but the male builds a decorated four-walled bower reinforced with up to 10 lb of stones.

The construction of bowers out of sticks and moss may have its origins in nest construction. Males of many bird species court females and induce them to ovulate by constructing true nests and offering food. It is possible that bowers began as courtship nests that became free of size constraints as the function of egg incubation was transferred to a separate nest built by the female alone. Decoration of bowers with coloured objects may have stemmed from courtship feeding, since displaying males of at least six species of

Simple maypole bower built by an elaborately ornamented male bowerbird, *Amblyornis macgregoriae* (mountains of New Guinea). The male has a striking, erectile orange crest three inches long. A circular moss platform with a rim is constructed about the base of a sapling, about which sticks are criss-crossed to form a tower. The bower has few or no decorations.



avenue builder, one maypole builder, and the mat builder *Archbodia papuensis sanfordi* pick up a decorative object in the bill and extend it towards the female. In some cases this decorative object is an edible fruit, although female bowerbirds have never been reported to take and eat the fruit. If courtship feeding was indeed the original meaning of this gesture, the meaning has been lost in the many cases where the display fruit has been replaced by an inedible object, such as a stone, shell, or human artefact like a cloth or marble.

Eight other species build bowers which could less easily have had their origins in nest construction and courtship feeding. The bowerbird *Scenopoeetes dentirostris* rakes an area of ground clean and decorates it with dozens of leaves carefully turned upside-down; the four birds of paradise (family Paradisaeidae, not Ptilonorhynchidae) of the genus *Parotia* rake an area of ground clean without decorating it; and the two birds of paradise of genus *Diphyllodes* rake an area of ground clean and strip the leaves off the trees over this area.

Does bower building imply that bowerbirds are intelligent and have an aesthetic sense? It certainly seems true, as Vellenga concludes, that many skills related to bower building and use have to be learned. Young males in green plumage spend about two years building rudimentary but increasingly complex bowers before acquiring blue adult plumage and building complete bowers. These 'practice bowers' are the joint efforts of several young males, which take turns placing and rearranging sticks, often clumsily and without success, occasionally

with the cooperation of more skilled older males. The young males do not paint these bowers and are less discriminating than adult males in choice of colour for bower decorations. The young males spend much time watching the displays, mating and other bower activities of adult males, and are displayed to by adult males.

As for aesthetic sense, colour preference in decorative objects varies with species and with individual, sometimes in a way that can be related to the bird's own colour. Adult males of *Ptilonorhynchus violaceus* prefer blue fruits, flowers, parrot feathers and man-made objects (see photograph on p.867 of ref. 10, depicting nine man-made objects recovered from bowers, all of them blue: a tooth-brush, toy airplane, pen, pencil-sharpener, clothes-pin, marble, shotgun cartridge, button and bottle-top). Blue is also the eye colour in both sexes, and is the colour of the adult male's plumage and bill. *Chlamydera calculata* lives in dry country, has brown plumage and eyes, selects pale ornaments and paints the bower brown. *Amblyornis flavifrons* displays with a blue fruit that affords maximum contrast with its golden crest, and consistently decorates the bower with fruits of three colours: the same blue fruit, green fruit that is the unripe form of the blue fruit, and a yellow fig similar to the colour of the abdomen and male crest. *Amblyornis inornatus* males vary individually: one used ornaments of seven different colours, another preferred brown and black, another favoured red and one specialized in white. The function of bower painting is mysterious, but a possible clue is that Vellenga twice observed a young male visiting a bower in the absence of its owner

and painting the walls with saliva, whereupon the owner returned, expelled the intruder and repainted the walls with green liverwort. Perhaps painting not only has an 'aesthetic' purpose but also provides an olfactory and visual mark of individual ownership, like the marking by urine or anal scent glands in many mammals.

Even if antecedents can be suggested for bower construction, decoration and painting, why have these display activities evolved to such an extreme degree in bowerbirds but in no other birds? One prerequisite is ample leisure time. Male bowerbirds spend most of their time repairing the bower, fighting rivals, raiding rivals' bowers and wooing females. Bowerbirds are among the largest frugivorous songbirds of Australia and New Guinea and are aggressive and dominant over other species because of their size. They may thus have preferential access to trees with concentrated fruit crops of high nutritional quality, enabling them to satisfy food needs in a short time and to devote remaining time to bower activities (see ref. 14). Other large tropical frugivores devote their leisure time to other bizarre social activities, such as the arboreal displays of most birds of paradise and the 'bowers' of the bird-of-paradise genera *Diphyllodes* and *Parotia* and the cotingid *Rupicola rupicola* (Cock-of-the-Rock).

It is apparent that major facets of bowerbird behaviour await further study. Gilliard argued that bowerbirds are lek species (ones with concentrated display grounds); this appears true for *Archboldia papuensis sanfordi* (known population confined to small areas on two New Guinea mountains), untrue for *Amblyornis flavifrons* (bowers spaced one-quarter of a mile apart) and unclear for other species. It is unknown whether simple, erratic, or nonexistent bower construction in the genera *Sericulus*, *Ailuroedus* and *Scenopoeetes* is a primitive condition or an advanced one: are these genera acquiring or losing the bower-building habit? Have bowerbirds co-evolved with certain tree species on whose fruit they specialize, as seems true for birds of paradise? Vellenga's work only begins to give tantalizing cues to the origin of colour preference, the origin of painting and the relative roles of instinct and learning in bower construction. To many ornithologists, bowerbirds remain the most intriguingly human of birds. □

The missing mass — now it's a gravitino!

from Joseph Silk

COSMOLOGISTS have a dire secret. They have not the slightest clue to the nature of the dominant mass constituent of the Universe — only the inherent implausibility of the idea prevents them from postulating the existence of innumerable copies of *Nature* floating in space. Preference has alternated for many years between two candidates: black holes and degenerate black dwarfs of the substellar variety. The mass density in the invisible form must exceed that in luminous material by one to two orders of magnitude.

The amounts of matter and luminosity in a given region are commonly measured in units relative to the mass and luminosity of the Sun. One finds a mass–luminosity ratio of about 2 in the solar vicinity; this means that the matter around us consists of stars typically of somewhat lower mass (luminosity being very sensitive to mass) than the Sun. In a great cluster of galaxies such as Coma, the ratio is about 300 (if one specifies that blue light is measured and that the Hubble constant is $50 \text{ km s}^{-1} \text{ Mpc}^{-1}$). Part of this difference can be understood in terms of the type of stars present in the cluster. Most of the cluster galaxies are ellipticals, which lack the young, hot, blue stars that contribute little to the mass of a galaxy such as ours but dominate its light. Thus, the mass–luminosity ratio of the underlying old disk and bulge populations of stars near the Sun that are most comparable with stars in ellipticals is about 6. Indeed, this is confirmed by measurements of the velocities of luminous K giants at a height of 1 kpc or more above the disk, which directly sample the local gravitational field. It follows that the total stellar content of the Coma cluster of galaxies amounts to only about 6/300 or 2 per cent of its dynamical mass.

Further, a substantial fraction (about 10 per cent) of the mass of the Coma Cluster has been discovered to be in the form of hot, X-ray-emitting gas. Thus, the total content of the Coma cluster that is in luminous form is only 12 per cent of its total mass. The remaining 88 per cent of the mass of Coma is invisible, at least hitherto, and its nature is a continuing source of speculation. A similar fraction of non-luminous matter appears to be present in the spiral-dominated groups of galaxies (although data on the intergalactic gas content of groups are very incomplete). These measurements both refer to scales of one or two million light years.

On smaller scales, galaxy rotation curves sample the distribution of dark matter in the outer parts of the galaxies. The predominance of flat rotation curves out to several times the optical radius of the galaxy indicates that the mass-to-light ratio is rising with increasing scale. The group

and cluster data suggest that the invisible mass fraction of the Universe continues to rise at least up to scales of millions of light years. Whether there is still more dark matter unclustered on larger scales is unknown. One can say that for the Universe to be of critical density, corresponding to closure, the required amount of dark matter has a mass–blue luminosity ratio of about 1,000. Measurements of the redshift–magnitude (or velocity–distance) relation for distant galaxies suggest that the mean density of the cosmos cannot much exceed this critical value. This ratio is only a factor of 3 larger than the Coma cluster value.

Does this signify that the Universe is near its closure density? Almost certainly not, since rich clusters are rather rare objects. Far more typical is the sparse group which has a characteristic mass–blue luminosity ratio of about 50. The large values of mass to luminosity in clusters are due in part to the older (and dimmer) star populations, and in part to the fact that a substantial amount of matter that would perhaps have formed spiral disks is in the intergalactic medium. The phenomenon that is responsible for the dark matter is a universal one. Even studies of galaxy haloes where little light is seen indicate that the ratio of mass to luminosity must locally be extremely large, amounting to several hundred or more. The evidence that exceedingly dark matter dominates the mean mass density of the Universe seems overwhelming. Such matter exists in galaxy clusters and groups, in the haloes of galaxies, and even in the solar vicinity, where, however, the principal contributor to the density is the old star population.

Particle physicists have joined the quest for dark matter in the cosmos. Experimental indications (still unconfirmed) of a finite rest mass for the neutrino have led to a new candidate for the dark matter. The big bang theory predicts that the number density of neutrinos at present should be comparable with the number of photons, both being relics of the primordial fireball that described the early evolution of the Universe. If the neutrino has a rest mass of 1 eV (or one-millionth of an electron mass), it would presently dominate the mass density of the Universe. With a neutrino rest mass as large as 100 eV, the Universe could even be of closure density. The existence of a massive neutrino has notable implications for the large-scale structure of the Universe. Neutrinos are collisionless particles and cannot sustain density fluctuations. As a consequence, all small-scale structure in the Universe would be suppressed, until

Joseph Silk is Professor of Astronomy at the University of California, Berkeley, California 94720.

1. Beccari *Ann. Mus. Civ. Genova* Ser. 1, 9, 382 (1877).
2. Beccari *Gardener's Chronicle*, 16 March, 332 (1978).
3. Marshall *Bower Birds: Their Displays and Breeding Cycles* (Clarendon, Oxford, 1954).
4. Gilliard *Birds of Paradise and Bower Birds* (Natural History Press, Garden City, New York, 1969).
5. Cooper & Forshaw *The Birds of Paradise and Bower Birds* (Collins, Sydney, 1977).
6. Vellenga *Emu* 80, 49 (1980).
7. Vellenga *Emu* 80, 87 (1980).
8. Vellenga *Aust. Bird Bander* 8, 3 (1970).
9. Chaffer *Aust. Zool.* 12, 295 (1959).
10. Green *Natn geog. Mag.* 152, 865 (1977).
11. LeCroy *Am. Mus. Novit.* 2714 (1981).
12. Diamond *Nature* 293, 257 (1981).
13. Gilliard *Scient. Am.* 209, 38 (1963).
14. Snow *The Web of Adaptation* (Collins, London, 1976).

very late times. This supports a theory of galaxy formation in which the galaxies arise from the fragmentation of large structures that have collapsed anisotropically into pancake-like sheets. Such anisotropies arise naturally in collapse when small-scale structure is suppressed, and lead to the formation of great voids between the clusters and superclusters of galaxies that develop in the high-density regions. One question that remains unanswered in such a theory concerns how the scale of a typical luminous galaxy is determined.

Now a new particle, the gravitino, has emerged from the creative minds of two particle theorists as a new candidate for the dark matter (H. Pagels and J. Primack *Phys. Rev. Lett.* **48**, 223;1982). The gravitino is a child of supersymmetry, associated with the search for a theory that unifies gravity along with the strong, weak and electromagnetic interactions. Supersymmetry unites fermions and bosons on a similar basis, and may help explain relations between certain large mass ratios involving the Planck mass, the mass associated with the grand unified theory symmetry breaking, and the mass associated with the electromagnetic-weak force symmetry breaking. Supersymmetry breaks down below energies of about 100 GeV, and results in the production of long-lived fermion partners of the graviton and the photon, the mediators of the only long-range forces. The new particles are called gravitinos and photinos, and both may have finite rest masses. The number of photinos is comparable with the number of neutrinos, and the implications of a finite photino mass are indistinguishable from those of a neutrino rest mass.

However, the gravitino uncouples sufficiently early that the predicted number of gravitinos is only about 10 per cent of the cosmic neutrino flux. Subsequent decouplings of other particles produce the bulk of the neutrinos. This leads to an interesting consequence: if gravitinos are of sufficient mass to account for the dark matter in the Universe, the gravitino rest mass must be some 10 times greater than the neutrino rest mass, or as large as 1,000 eV if gravitinos are at closure density. As the Universe expands, the more massive gravitinos become non-relativistic at an earlier epoch than do the neutrinos. The horizon scale at this instant provides a measure of the minimum primordial fluctuation scale that can survive the free streaming of particles at the speed of light. For neutrinos of mass 30 eV, the perturbation amplitude peaks at $4 \times 10^{15} M_{\odot}$, whereas for 1 keV gravitinos the perturbation spectrum peaks at about $10^{12} M_{\odot}$. This is just the mass-scale associated with the dark haloes of galaxies. If we take the coincidence seriously, then gravitinos of mass 1 keV are the dominant source of matter in the Universe, and may account for the characteristic scale of galaxies.

Indeed, it turns out, according to

unpublished calculations by J. R. Bond, A. S. Szalay and M. S. Turner, that gravitinos may lead to the best of all possible scenarios for large-scale evolution. For their mass spectrum extends all the way from scales of supercluster mass down to scales comparable to that of dwarf galaxies. Unlike the neutrinos, gravitinos do not dominate the mass density of the Universe until long after they have become non-relativistic, and this enhances the growth of large-scale gravitino perturbations relative to smaller-scale perturba-

tions, whose growth is inhibited by the inertia of the radiation-dominated mass content of the Universe.

Dark matter may therefore consist of massive neutrinos or gravitinos, or perhaps some more exotic species of elementary particle whose presence arises from unification theories that describe the first 10^{-36} second of the Universe. Such cosmions are the bane of the astronomer's life, for they are likely to be forever invisible, yet they provide a potential solution to some of his greatest problems.

The regulation of contraction in cardiac muscle

from D.A. Eisner and D.G. Allen

ONE hundred years ago Sydney Ringer, a physician working at University College London, discovered that calcium is required for the heart to contract. Only in the past twenty years, however, have the intracellular mechanisms responsible for this calcium dependence become clear and the central role of intracellular calcium in the control of cellular processes become widely appreciated. In the past five years our understanding of the regulation of cardiac contraction has taken a leap forward with the introduction of techniques for measuring the crucial intracellular ion activities (Ca^{2+} , Na^{+} , H^{+} , K^{+} and Cl^{-}) and the development of biochemical methods for assessing the regulation of protein function. The field of cardiac electrophysiology also seems poised for a period of rapid growth with the recent introduction of single cell and patch clamping techniques. Each of these fields was represented at a recent meeting* at University College London which concentrated on two aspects of the regulation of cardiac contraction. The first was that of how the low resting level of calcium is maintained and how the increase of intracellular calcium required for contraction is produced. The second concerned mechanisms other than changes of calcium concentration which may be involved in the regulation of cardiac contraction.

The role of $\text{Na}^{+}/\text{Ca}^{2+}$ exchange in the regulation of intracellular calcium dominated the first half of the meeting. L.J. Mullins (University of Baltimore) suggested that the intracellular calcium concentration is controlled by a voltage-dependent $\text{Na}^{+}/\text{Ca}^{2+}$ exchange which, at rest, pumps calcium ions out of the cell in exchange for sodium ions that enter by flowing down their electrochemical gradient. A novel aspect of his model is that it attributes much of the calcium entry during the long cardiac action potential to

this mechanism. It was suggested, by analogy with experiments on squid axons, that depolarization makes the $\text{Na}^{+}/\text{Ca}^{2+}$ exchange bring Ca^{2+} into the cell and extrude Na^{+} . In contrast, H. Reuter (University of Berne) stressed the importance of conventional Ca^{2+} channels in providing the influx of calcium ions during the action potential. Recent recordings from single channels have removed lingering doubts about the existence of the calcium channel in the heart. It is clear that quantitative measurements are now required to establish the relative contributions of Ca^{2+} channels and $\text{Na}^{+}/\text{Ca}^{2+}$ exchange to the rise of Ca^{2+} which produces contraction.

Papers given by D. Eisner (University College London) and by D. Ellis (University of Edinburgh) were concerned with altering the magnitude of $\text{Na}^{+}/\text{Ca}^{2+}$ exchange and examining the effects of contraction. It is clear that small changes of internal sodium concentration have enormous effects on contraction. These effects depend on the membrane potential and the external sodium concentration in a way which is, at least, consistent with the predictions of a voltage-dependent $\text{Na}^{+}/\text{Ca}^{2+}$ exchange model. However, it was also shown that changes of internal and external sodium concentrations have pronounced effects on intracellular pH and these may also affect contraction.

Evidence that the contractility of cardiac muscle can be regulated by mechanisms other than changes in the calcium levels was also presented. Direct evidence was presented by D. Allen (University College London) who, using the bioluminescent protein aequorin as a calcium indicator, showed that as well as increasing the cytoplasmic calcium concentration during contraction, adrenaline reduces the sensitivity

*The Physiological Society Symposium on 'The Regulation of Contraction in Cardiac Muscle' was held on 25-26 March 1982.

D.A. Eisner and D.G. Allen are in the Department of Physiology, University College London, Gower Street, London WC1E 6BT.

of the contractile proteins to calcium. In an examination of the biochemical basis of this phenomenon, P. England (University of Bristol) showed that the application of adrenaline results in a protein kinase-dependent phosphorylation of troponin and other proteins. The phosphorylation allows the troponin to speed up its release of bound calcium and may help heart muscle to relax more quickly — of obvious importance for the high heart beat rates produced by adrenaline. Other proteins which become phosphorylated in the presence of adrenaline include a component of the sarcoplasmic reticulum calcium pump and a surface membrane protein. These phosphorylations may therefore affect movements of calcium between the cytoplasm and the sarcoplasmic reticulum and between the cytoplasm and the extracellular space. It was suggested by F. Flitney (University of St Andrews) that some of these phosphorylations could be controlled by the relative levels of cyclic AMP and cyclic GMP.

A major difficulty inherent in studies of the regulation of contraction is that the

intracellular calcium concentration cannot be controlled when the studies are performed on intact cells. Unfortunately, biochemical studies suffer from the disadvantage that cell constituents involved in the regulation of contraction may be lost. To circumvent this problem, S. Winegrad (University of Philadelphia) has attempted to make cells which are 'hyperpermeable' to calcium ions but which retain larger substances within the cell. Using this technique he produced evidence for the control of the calcium sensitivity of contractile proteins by phosphorylation of troponin. This approach came under attack from D. Miller (University of Glasgow) who suggested that such cells are not really permeable to calcium ions. It appears that this problem must be resolved before the results can be unequivocally interpreted. Winegrad also described recent experiments which suggest that a proportion of the myosin molecules are normally inactivated but can be activated by a cyclic AMP-dependent process. If confirmed, this raises the prospect of an additional mechanism whereby force production in cardiac muscle can be varied. □

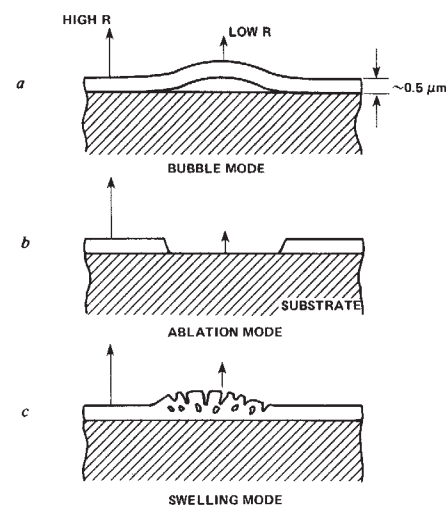


Fig. 2 Recording mechanisms in hydrogenated amorphous semiconductor films. a, Bubble formation; b, ablation; c, sponge formation

by plasma etching the semiconductor (Si or Ge) layer in the presence of aluminium. Atoms are sputtered from the aluminium surface and deposited on the semiconductor film to form a relatively thin and discontinuous layer of aluminium islands with dimensions of 100 nm or less. These act as an etching 'mask' that produces a microscopic needle-like texture on the etched semiconductor surface. The texture has a relatively broad-band anti-reflection characteristic (previously used in solar thermal collectors¹¹) and can be locally melted by an efficiently coupled recording beam. Surface tension promptly restores a more nearly planar geometry to the melted area (see Fig. 1) and the data so recorded may be retrieved by detecting the increased reflectance of the smoothed region. Some critical problems of this approach do remain — it is not yet known how much playback signal-to-noise ratio (SNR) is reduced by the inherent 'graininess' of the textured surface, and whether the plasma etching can be sufficiently well controlled to act uniformly over disk-sized areas ($\sim 0.1 \text{ m}^2$).

Bösch¹², also of BTL, has investigated the recording characteristics of amorphous hydrogenated semiconductors prepared by reactive sputter deposition of the semiconductor material in a gas mixture containing 25 per cent hydrogen in argon at 10 mtorr. Three distinct recording mechanisms are possible, all of which are activated by the effusion of the hydrogen from the region heated by the laser beam; a process which occurs at a relatively low temperature and is strongly dependent on the choice of semiconductor. At low powers a blister or bubble region (Fig. 2a) forms in amorphous hydrogenated silicon semiconductors. At higher powers, ablation of the films occurs, apparently without melting, since no re-solidified residue is observed at the rim of the pit (Fig. 2b). In the case of amorphous hydrogenated germanium semicon-

Recent developments in optical storage technology

from Alan E. Bell

RESEARCH and development in the field of high-density optical data storage media and devices continue to move forward at a brisk pace — in the eighteen months since last reviewed in these columns¹, several companies have committed themselves to development efforts and the introduction of commercial optical disk computer memories can be expected in the next few years. Such devices will offer rapid random access ($\sim 200 \text{ ms}$) to a database of up to 10 gigabytes per disk at a cost of 10^{-6} cents per byte or less and seem likely to leap-frog the projected improvements in the currently ubiquitous magnetic data storage media. The impact in many areas of electronic data processing, ranging from the 'electronic office' to large data-base archives and even to small business and personal computer systems, will be profound².

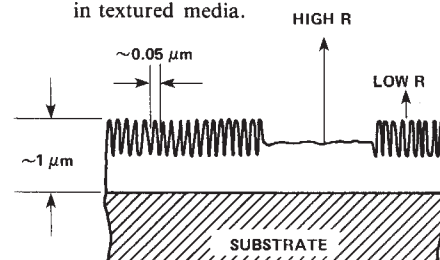
The two most critical components of the optical memory system are the diode laser read/write light source and the optical storage medium itself. The diode laser source is much preferred to the alternative gas laser sources, mainly because of the small size and power requirement, the capability for direct modulation of the optical record beam by the input data signal and the potentially high reliability

and lifetime. Recent advances in the design and fabrication³⁻⁸ of high-power ($>40 \text{ mW}$ output at 10 MHz modulation) diode lasers have made them available as commercial products, albeit at an initially high cost. With manufacturing scale-up and improved process yields, the cost should rapidly decrease to less than \$100 (ref. 9).

The 'ideal' optical storage medium has yet to be found and the past year has seen reports of many new recording media and configurations. Several new write-once (that is, irreversible) recording mechanisms have been identified and significant gains in the optimization of reusable (that is, reversible) media have been made.

Workers at Bell Telephone Laboratories (BTL) have reported the recording characteristics of irreversible media, microscopically textured low-reflectivity films of vacuum-deposited semiconductors¹⁰. The textured surface layer (see Fig. 1) is formed

Fig. 1 Recording mechanisms in textured media.



Alan E. Bell is Senior Staff Physicist at Exxon Research and Engineering Company, PO Box 45, Linden, New Jersey 07036.

ductors, the effusion of hydrogen produces a micrometre-sized porous region, somewhat resembling a sponge (Fig. 2c). The hydrogenated semiconductors provide a very stable storage medium but recording with a diode laser source is difficult because the wavelength of the beam (~ 820 nm) is in the region of the bandgap energy and absorption is therefore low.

Bubble formation is perhaps the most intriguing recording mechanism reported by Bösch and has previously been described by workers at 3M¹³ and Thomson-CSF¹⁴. In the Thomson-CSF approach, micrometre-sized blisters are formed in a thin layer of a noble metal alloy (for example, Au or Pt alloy) by partial vaporization of an underlying polymer layer. The scattering of the read beam by the blister results in a relatively low effective reflectivity of the blister compared with the undisturbed planar regions (Fig. 3a). Workers at 3M and others¹⁵ have combined the bubble mode storage medium with the anti-reflection trilayer structure^{1,16} (Fig. 3b). The trilayer increases optical efficiency ($A \leq 0.9$) and thus improves sensitivity. When the blister forms in the uppermost (absorber) layer the interference properties which cause the structure to have low reflectivity are unbalanced and a high-reflectivity blister region forms. One advantage of the blister approach is its potential for extremely high SNR. Irregularly shaped residue at the pit rim, often the limiting factor in the playback SNR of melted pit media, are absent. Furthermore, since the recording sensitivity of bubble mode media is dictated by the thermal vaporization properties of the underlying polymer layer, rather than those of the bubble-forming absorber layer, relatively refractory and stable materials can be used for the latter without sacrificing the recording sensitivity normally associated only with low melting point, often unstable, recording materials¹⁷.

A novel approach to optical storage media has been disclosed by workers at IBM¹⁸. A layer of amorphous silicon is

coated with a noble metal layer, such as Rh or Pt, to a thickness which has a low reflectivity at the recording wavelength. Heat generated by the absorption of the recording beam energy causes the two layers to react and form a microscopic mark consisting of relatively highly reflective noble metal silicide (Fig. 4). This approach provides archival stability but appears to be relatively insensitive. Unfortunately, if an attempt to improve sensitivity is made by choosing metals (such as Pd) which form silicides at relatively low temperatures, then the rate of reaction between the layers, even under room temperature storage conditions, leads to an unacceptably low archival lifetime for the disk.

For a number of data storage applications, for example the filing of legal, financial and medical records, the write-once storage media discussed above provide a positive guarantee that the records have remained unchanged since input. Nevertheless, computer-operating systems have evolved for a period of 30 years with data storage devices that offer erasure and rewrite capability. Two potential candidates for reversible optical storage media that have been intensively investigated are the chalcogenide alloys such as TeGeAs¹⁹ and magneto-optic alloys such as MnBi²⁰ and GdCo²¹.

The chalcogenide alloys exist at room temperature in either the crystalline or amorphous form, each having different optical properties. An amorphous mark is recorded in a polycrystalline film by briefly melting (~ 50 ns) the crystalline material, and the subsequent rapid cooling ($\sim 10^{10}$ s⁻¹) 'freezes' the random atomic structure of the melt to produce an amorphous solid. Recrystallization (that is, erasure) can be accomplished by laser beam heating the amorphous region to a temperature above the glass transition temperature but below the melting point. If the temperature is maintained for longer than about 1 μ s, the amorphous dot is annealed and so returned to the original polycrystalline state. Bell and Spong²² demonstrated reversible optical recording of video signals with an encapsulated tellurium trilayer disk. The encapsulation inhibits the irreversible process of hole formation during the molten phase of the amorphization step. Despite earlier reports²³ indicating limited cyclability of chalcogenide alloys it was found that with pure tellurium up to fifty record and erase cycles were executed on a single track without significant deterioration of signal quality.

In contrast to the chalcogenide alloy phase-change materials, the number of cycles of magnetization reversal possible in conventional magnetic media appears to be unlimited. In reversible magneto-optic materials, magnetization reversal is accomplished by Curie-point writing or other techniques²⁴. In Curie-point writing the magnetization of a micrometre-sized region of the film is reversed by first using the record

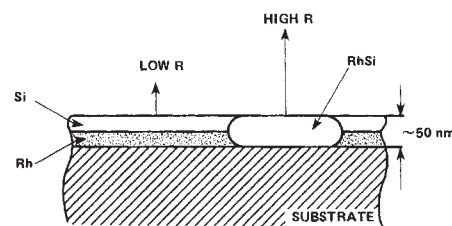


Fig. 4 Alloying bilayer approach.

laser beam to heat temporarily the region above the Curie temperature and then cooling in the presence of an external magnetic field oriented antiparallel to the initial (perpendicular to the substrate) magnetization direction. The presence of the magnetization reversal is detected by the sign of the Kerr or Faraday polarization rotation produced in light respectively reflected from or transmitted by a magnetized layer. In the past, playback SNR has been limited by the low optical signal contrast between oppositely magnetized regions. It has been known for some time²⁵, however, that the Kerr effect, in particular, can be increased by coating the magneto-optic material within an anti-reflection structure¹⁶. Mansuripur *et al.*²⁶ showed that the incorporation of a magneto-optic material into the trilayer structure can lead to a significant increase in signal contrast and thus SNR on playback. The incorporation of recently developed ternary amorphous magneto-optic alloys such as TbGdFe²⁷ into trilayer structures promises the practical realization of reversible high-density optical data storage. Incidentally, the interference properties of the trilayer structure can also yield significant increases in SNR of both Raman and Kerr spectroscopy used in basic studies^{28,29} of the properties of very thin films (≤ 10 nm).

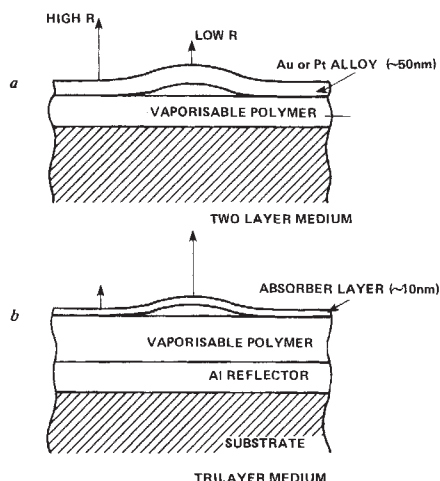


Fig. 3 Bubble mode recording media. a, Bubble formation in single layers; b, bubble formation in trilayer structures.

1. Bell, A.E. *Nature* **287**, 583 (1980).
2. Goldstein, C.M. *Science* **215**, 862 (1982).
3. Ueno, M. *IEEE J. Quantum Electron.* **QE-17**, 2113 (1981).
4. Eitenberg, M. & Botez, D. *Electron. Lett.* **18**, 153 (1982).
5. Botez, D., Channin, D.J. & Eitenberg, M. *SPIE Proc.* **321** (in the press).
6. Kawabe, H. *et al.* 13th Jap. Conf. on Solid State Devices (1981).
7. Saito, K. & Ito, R. *IEEE J. Quantum Electron.* **QE-16**, 205 (1980).
8. Botez, D. *IEEE J. Quantum Electron.* **QE-17**, 2290 (1981).
9. Gamo, H. *Laser Focus* **18**, 57 (1982).
10. Craighead, H.G. & Howard, R.E. *Appl. Phys. Lett.* **39**, 532 (1981).
11. Stephens, R.B. & Cody, G.D. *Solar Energy Mater.* **1**, 397 (1979).
12. Bösch, M.A. *Appl. Phys. Lett.* **40**, 8 (1982).
13. Robbins, W. *et al.* Conf. on Lasers and Electro-Optics CLEO'81 Tech. Dig. 122, Washington DC (1981).
14. Cornet, J. *et al.* Conf. on Lasers and Electro-Optics CLEO'81 Tech. Dig. 122, Washington DC (1981).
15. Bell, A.E. *J. appl. Phys.* (in the press).
16. Bell, A.E. & Spong, F.W. *IEEE J. Quantum Electron.* **QE-14**, 487 (1978).
17. Lou, D.Y. *et al.* *J. Vac. Sci. Technol.* **18**, 78 (1981).
18. Tu, K.N. *et al.* *Appl. Phys. Lett.* **39**, 927 (1981).
19. Von Gutfeld, R.J. & Chaudhari, P. *J. appl. Phys.* **43**, 4688 (1972).
20. Chen, D. *et al.* *IEEE Trans. Mag.* **MAG-9**, 66 (1973).
21. Chaudhari, P. *et al.* *Appl. Phys. Lett.* **22**, 337 (1973).
22. Bell, A.E. & Spong, F.W. *Appl. Phys. Lett.* **38** 920 (1981).
23. Smith, A.W. *Appl. Opt.* **13**, 765 (1974).
24. Chen, D. *Appl. Opt.* **13**, 767 (1974).
25. Lissberger, P.H. *J. opt. Soc. Am.* **51**, 957 (1961).
26. Mansuripur, M. *et al.* *SPIE Proc.* **321** (in the press).
27. Imamura, N. & Ota, C. Conf. on Lasers and Electro-Optics CLEO'81 Tech. Dig. 106, Washington DC (1981).
28. Connell, G.A.N. *et al.* *Appl. Phys. Lett.* **36**, 31 (1980).
29. Connell, G.A.N. *Appl. Phys. Lett.* **40**, 212 (1982).

Optical data storage technology is at a particularly exciting phase in its development. Steady improvements in both key areas of semiconductor laser sources as well as the optical storage medium promise the realization of optical storage memory devices within the next few years. Much further research into the

microscopic degradation and ageing mechanism in thin films will be required to develop a truly archival (>100 yr) storage medium; and the development of independently addressable semiconductor laser arrays that will permit extremely high (10–100 Mbytes s⁻¹) record and playback data rates. □

Rival transmitters in visual transduction

from Helen Saibil

MANY recent papers suggest that cyclic GMP plays a major role in visual transduction, relaying the light stimulus from the visual pigment, rhodopsin, to the sodium channels of the retinal rod. Until a few years ago it was thought that this role was exclusively played by calcium ions, but the new results suggest that calcium and cyclic GMP interact to regulate the sodium conductance.

Absorption of light by rhodopsin, the visual pigment in the rod outer segment (ROS) disc membranes, changes the concentration of a cytoplasmic transmitter that controls the conductance of sodium channels in the ROS plasma membrane. These channels (or carriers) allow a continual sodium influx, or dark current, which is reduced by the absorption of even a single photon of light. It is this reduction (rather than an increase) of 'dark current' that brings about the neural response to a visual stimulus that is transmitted through the retina to the brain. It was originally proposed that bleached rhodopsin released stored Ca²⁺ from the discs because it was known that Ca²⁺ blocks the sodium conductance (see refs in ref 1). The recent discovery of a large and rapid light-induced release of Ca²⁺ from rods supports the idea that light causes the required rise in cytoplasmic free Ca²⁺ (refs 1,2). But recent papers on the sequence and kinetics of a series of enzyme reactions in ROS provide evidence for another mechanism in which rhodopsin bleaching controls the concentration of cGMP.

As reported in these pages about two years ago³, rhodopsin bleaching can rapidly decrease cGMP concentration in ROS (see refs 4-6). In contrast to Ca²⁺, injection of cGMP depolarises the rods, and delays the subsequent photoresponse⁷. Ca²⁺ could thus act as a positive transmitter, absent (or bound) in the dark and released by light to block channels⁸ and cGMP could act as a negative transmitter keeping channels open in the dark, with light reducing its concentration and allowing channels to close. As well as having opposite effects on the sodium channels, Ca²⁺ and cGMP also have

antagonistic effects on some phosphorylation levels in ROS⁹. Various combinations of Ca²⁺ and cGMP have been proposed to function as intracellular transmitter; these are reviewed in a recent collection of articles on the subject¹⁰ which raises the exciting prospect that electrophysiological and biochemical approaches may be brought together from opposite ends of the transduction process.

The central question is how photoexcited rhodopsin controls the sodium conductance. From the outside of the rod, electrophysiologists have sought to characterize the sodium channels and the process controlling their conductance. From the inside, biochemists have looked for reactions activated by photoexcited rhodopsin (R*).

The first approach, using intact, functioning rods, has yielded precise information on the electrical response, showing that the transduction process has four major delay stages and that a single photoisomerization causes many channel closures (summarized in ref 11). Electrical recordings also establish that excitation and light adaptation are both mediated by diffusible transmitter substances in the outer segment¹²⁻¹⁴. They do not, however, provide information on the chemical nature of the controlling process, that is, the link between R* and transmitter concentration. To study this, biochemists have isolated and broken open outer segments to get at the cytoplasm where R* activation of the transduction takes place. This has led to the discovery of an enzyme cascade which links rhodopsin bleaching to cGMP hydrolysis in the ROS cytoplasm, a chain of reactions with close parallels to hormone receptor-adenylate cyclase activation^{4,15}.

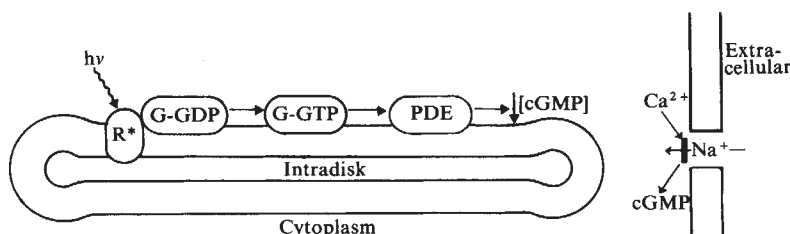
Rhodopsin bleaching stimulates this cascade, as shown in Figure 1. R* interacts with a three-subunit GTP-binding protein (called G-protein, GTPase or even transducin¹⁵) and catalyses its exchange of bound GDP for GTP. A single R* molecule can catalyse many such exchanges sequentially. Each G-GTP in turn activates many cGMP phosphodiesterases (PDE) which hydrolyse cGMP to 5'GMP. G and PDE, as well as a rhodopsin kinase, are bound to the cytoplasmic surface of discs¹⁶. G slowly hydrolyses bound GTP back to GDP, an activity which earned its original name GTPase⁴, switching off PDE activation. The rhodopsin kinase may deactivate R* by ATP-dependent phosphorylation (ref. 6, but see ref. 9).

Changes in the far-red light scattering of ROS suspensions that were found to be elicited by a brief bleaching flash¹⁷⁻¹⁹ are now being used to follow the kinetics of the earlier parts of the enzyme cascade. Kühn's key discovery that the binding of G-protein to the disc surface depends on ionic strength, light and GTP has made it possible to separate G from unbleached discs so that light scattering signals can be directly related to the amount of G present in reconstituted disc+G mixtures. Without GTP there is a strong light-induced binding of G to the discs.

Kühn, in collaboration with Bennett, Michel-Villaz and Chabre²⁰, went on to show that the light-induced drop in transmission (increase in turbidity) of disc suspensions in the absence of GTP reflects the formation of a 1:1 R*-G complex (fig 2). This 'binding signal' saturates when all the available G is bound (10 per cent of rhodopsin in native ROS). If > 20 μM of GTP is present, an increase in transmission, the so-called 'dissociation signal', occurs as the G-GTP dissociates from R*. This signal is seen only when GDP-GTP exchange occurs and it saturates at <0.8 per cent bleaching. Moving on to the next stage in the cascade, PDE activation, Bennett²¹ used light-scattering techniques in parallel with the proton release method of Liebman and Pugh⁶ to follow cGMP hydrolysis. It begins soon after the onset of the dissociation signal (Fig 2) and has the same dependence on nucleotides and light intensity, showing that PDE activation by G-GTP¹⁵ is rapid.

The rod electrical response has a characteristic s-shaped delay which shortens with

Fig.1 Photoexcited rhodopsin (R*) binds G-protein and catalyses its exchange of bound GDP for GTP. G-GTP dissociates from R* and activates the phosphodiesterase (PDE) which hydrolyses cyclic GMP. The sodium conductance is decreased by Ca²⁺ and increased by cyclic GMP.



Helen Saibil is in the Biophysics Department, University of London King's College, Drury Lane, London WC2R 2LS.

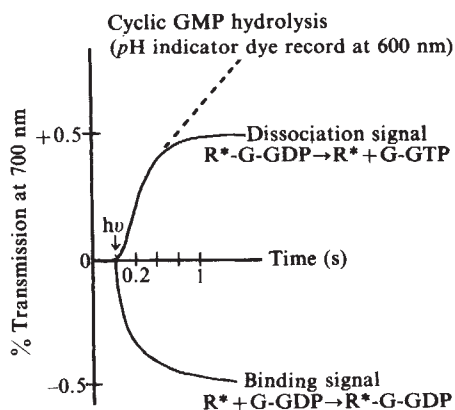


Fig. 2 Far-red light-scattering signals induced by rhodopsin bleaching in a suspension of ROS fragments or from washed discs reconstituted with isolated enzymes. The binding signal indicates the formation of a stable 1:1 complex between R^* and G. The dissociation signal results from GDP-GTP exchange on G. Hydrolysis of cyclic GMP to 5'-GMP + H^+ is monitored by a pH-sensitive dye.

increasing light intensity. It is intriguing that the dissociation signal shows very similar behaviour, but it is dangerous to compare experiments on completely intact rods to those on completely disrupted ones. A possible explanation for the delay, advanced by Liebman and Pugh⁶, stems from the fact that rhodopsin molecules are more abundant than the proteins with which they interact. As a result, at low bleaches when there are fewer R^* created it takes longer for an isolated R^* to find its target protein — a process dependent on lateral diffusion. Lateral diffusion seems rate-limiting in the analogous hormone receptor system where chemically increasing the membrane fluidity speeds up

cyclase activation²².

Both Ca^{2+} and cGMP have drawbacks as candidates for the intracellular transmitter. From the biochemical viewpoint, the persistent failure to relate Ca^{2+} uptake or release to changes in rhodopsin or disc membranes²³ and the relative completeness of the R^* -cGMP pathway makes cGMP an attractive alternative. From the electrophysiological viewpoint, cGMP hydrolysis has kinetics comparable to the electrical

response only if the hydrolysis is triggered by light levels much higher than needed to trigger the corresponding electrical response (see refs in refs 5 & 11), and there are questions about its time course *in vivo*²⁴. We have had already had to accept that the minor proteins in ROS cannot be dismissed as irrelevant to visual excitation; now we may have to learn to live with multiple regulation of the sodium conductance by calcium and cGMP. □

Scrapie agent: prions or virinos?

from Richard H. Kimberlin

In the last few weeks considerable publicity has been given to a claim, by Stanley B. Prusiner of the University of California, San Francisco, that the agent causing scrapie — a degenerative disease of the central nervous system in sheep and goats — is a novel kind of proteinaceous infectious particle¹. The particle, named the 'prion' by Prusiner, is claimed to be either a very small oligonucleotide surrounded by a tightly packed protein or, more remarkably, an infectious protein that contains no nucleic acid at all. This latter possibility is, as Prusiner says, 'clearly heretical' and raises all the problems of how a non-nucleic acid-containing particle can replicate itself. Have we indeed reached the point where the existence of an infectious protein seems likely? To answer this question Prusiner's work needs to be set in the context of 20 years' research in a very difficult area.

In the early 1960s, scrapie was experimentally transmitted from its natural hosts, sheep and goats, to mice and other rodent species, allowing a wide variety of experimental models of scrapie, novel techniques for identifying different strains of agent and a conceptual framework for the genetic control and slow pathogenesis of the disease. Equally important, quantitative studies on the nature of the transmissible agent could begin using titration in mice to measure infectivity (see ref. 2 for a review).

Since then, various classes of microbial agents, such as bacteria, mycoplasmas and viroids, have been excluded as the scrapie agent. The agent has many of the biological properties of a virus but is 'unconventional' in its high degree of physicochemical stability and apparent lack of specific antigenicity. However, we still have very little detailed knowledge of the nature of the scrapie agent.

There are two main reasons for this slow progress. First, the infectious agent is notoriously 'sticky': it seems to aggregate to cell components and has not been sufficiently purified to make biochemical characterization easy. Consequently, its physicochemical properties have been

determined only on crude or partially purified extracts. Second, titration in animals is still the only method of measuring the infectious agent. Even with the quickest models of scrapie in mice and hamsters, a complete titration takes 200–300 days. There are also difficulties in interpreting the results: infectivity titres obviously reflect the amount of scrapie agent in an inoculum but they are also influenced by the efficiency of the infection process. This in turn depends on factors which may include varying host constituents and residual reagents in inocula, and the degree of dispersion or aggregation of a 'sticky' agent after treatment. A change in scrapie titre may be particularly difficult to interpret if a given reagent interacts with more than one molecular species, for example, protein and nucleic acid.

Prusiner has worked largely with a short-incubation hamster model of scrapie and has added a sizeable amount of evidence, supporting earlier observations that scrapie infectivity can be reduced by treatments which denature, disaggregate or breakdown proteins^{3–6}. The loss of infectivity is generally quite large and it seems reasonable to conclude that the infectivity of scrapie agent involves protein.

However, the unresolved issue is whether such protein is 'contained' in the agent as an integral part of its structure or merely associated with it. This leads to the key question: is there a scrapie-specific protein? Neither Prusiner nor anyone else has any published data showing that there is, but, for the moment, let us assume that one does exist. Conventional thinking would then be that the protein is encoded by a specific nucleic acid which forms the genome of the agent. However, Prusiner has produced two interesting lines of evidence which he regards as indicating the absence of a scrapie nucleic acid.

First, treatments known to inactivate nucleic acids, even though not specific for

Richard H. Kimberlin is Head of the Pathogenesis Section, ARC & MRC Neuropathogenesis Unit, Edinburgh.

1. Yoshikami, S., George, J.S. & Hagins, W.A. *Nature* **286**, 395 (1980).
2. Gold, G.H. & Korenbrot, J.I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5557 (1980).
3. Fatt, P. *Nature* **280**, 355 (1979).
4. Bitensky, M.W., Wheeler, G.L., Yamazaki, A., Rasenick, M.M. & Stein, P.J. in ref. 10, p.238.
5. Bownds, M.D. in ref. 10, p.203.
6. Liebman, P.A. & Pugh, E.N. in ref. 10, p.157.
7. Miller, W.H. & Nicol, G.D. in ref. 10, p.417.
8. Yau, K.-W., McNaughton, P.A. & Hodgkin, A.L. *Nature* **292**, 502 (1981).
9. Hermolin, J., Karell, M.A., Hamm, H.E. & Bownds, M.D. *J. gen. Physiol.* **79**, 633 (1982).
10. Miller, W.H. (ed.) *Curr. Topics in Membranes and Transport* **15** (1981).
11. Matthews, G. & Baylor, D.A. in ref. 10, p.4.
12. Donner, K.O. & Hemila, S. *Med. Biol.* **56**, 52 (1978).
13. Bastian, B.L. & Fain, G.L. in ref. 10, p.341.
14. Lamb, T.D., McNaughton, P.A. & Yau, K.-W. *J. Physiol., Lond.* **319**, 463 (1981).
15. Fung, B.K.-K., Hurley, J.B. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 152 (1981).
16. Kühn, H. in ref. 10, p.172.
17. Asai, H., Chiba, T., Kimura, S. & Takagi, M. *Expl. Eye Res.* **21**, 259 (1975).
18. Bignetti, E., Cavaggoni, A., Fasella, P., Ottonello, S. & Rossi, G.L. *Molec. cell. Biochem.* **30**, 93 (1980).
19. Reichert, J. & Hofmann, K.P. *Biophys. struct. Mech.* **8**, 95 (1981).
20. Kühn, H., Bennett, N., Michel-Villaz, M. & Chabre, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6873 (1981).
21. Bennett, N. *Eur. J. Biochem.* **123**, 133 (1982).
22. Hanski, E., Rimon, G. & Levitzki, A. *Biochemistry* **18**, 846 (1979).
23. Szuts, E. in ref. 10, p.291.
24. Ebrey, T.G., Kilbride, P., Hurley, J.B., Calhoun, R. & Tsuda, M. in ref. 10, p.121.

them, are not found to destroy scrapie infectivity. Infectivity is relatively resistant to heating at 65°C in the presence of zinc ions, to photochemical inactivation after treatment with psoralens and to treatment with 0.5 M hydroxylamine¹. Failure to inactivate may, however, simply mean that these reagents did not gain access to the putative nucleic acid, perhaps because of protection by protein forming part of the infectious agent or by contaminating host molecules. In any case, as Prusiner points out, not all viruses are inactivated by psoralens or hydroxylamine.

Second, in preliminary studies using detergents and gel filtration, some scrapie infectivity eluted between bovine serum albumin and ovalbumin markers¹. Assuming a globular structure for the infectious agent, and recognizing the potential problems of determining molecular weights in this way, Prusiner argues that scrapie agent may have a molecular weight of 50,000 or less. He considers that a proteinaceous agent of this size would be too small to contain a nucleic acid of more than just a few nucleotides.

Interesting though they are, these two findings are not by themselves compelling reasons for considering highly unorthodox models of scrapie agent, particularly when faced by some of the problems these pose. Prusiner suggests three such models.

In the first model, the agent contains no nucleic acid and scrapie-specific protein acts as its own template, either directly, or by 'reverse translation'. It is difficult to conceive how proteins could code for their own synthesis although scrapie models of this kind were proposed as long ago as 1967 (ref. 7). But since then the requirements of a scrapie model have become more stringent; the copying process must have a precision compatible with the existence of several different, genetically stable strains of scrapie⁸.

In Prusiner's second model, the agent

contains no nucleic acid but has protein which induces host genes to code for its own synthesis. The third proposes that there is a very small scrapie oligonucleotide which does not code for protein but acts as a regulatory element promoting the synthesis of agent. In these models, therefore, it is the host nucleic acid that encodes scrapie-specific information. This is certainly conceivable but difficult to reconcile with evidence that scrapie agents behave like pathogens with their own coded information: the natural disease is infectious⁹, various strains of agent can be transmitted experimentally to many different species⁸, some strains give rise to mutants¹⁰ and strain selection can occur on serial passage of mixtures¹¹. To explain these properties in terms of host coding introduces a complexity that seems unnecessary in the absence of data which necessitate it.

A much simpler working hypothesis which fits both established facts and even Prusiner's recent data may be proposed. It makes only two assumptions: first, that there is indeed a scrapie-specific nucleic acid, as indicated by the properties of replication, strain variation and mutation; and second, that, as in viroids, the scrapie nucleic acid is not translated¹².

The hypothesis has some interesting consequences. The scrapie nucleic acid could be very small as all it has to do is be replicated (using host enzymes) and interact in strain-specific ways with the host to produce disease. The latter it could do by acting as, or coding for, some kind of regulatory nucleic acid. Alternatively, disease could be a consequence of the

binding of the scrapie-specific nucleic acid to host protein needed to form an infectious agent, especially if the protein normally has an important biological function. Since the protein associated with the agent is host derived we have a very simple explanation for the apparent absence of specific antigenicity.

This hypothesis may also explain our current failure to purify agent or to identify it in the electron microscope: if the scrapie-specific nucleic acid is covered with host protein it may well be 'sticky' and hard to purify from other host proteins and, unless it has a regular, virus-like nucleoprotein structure, it could be hard to recognize. An agent of this kind would be novel but fit nicely into a biological niche lying between viruses, which specify some of their own proteins, and viroids which need no protein at all. The neologism, 'virino', has already been suggested for a novel agent of this kind¹³, and may be preferable to Prusiner's term 'prion' because the latter emphasizes a molecular species which may not be the most important one, as did the conclusion in 1935 that tobacco mosaic virus was proteinaceous¹⁴.

However, the main point of these speculations is that we do not yet need to build hypotheses outside the current framework of molecular biology to accommodate the scrapie agent. The real need is to assemble more hard facts and, just as Prusiner¹ says, this means "purification of the scrapie agent to homogeneity and the subsequent identification of its macromolecular components"; undoubtedly a difficult task but not an impossible one. □



100 years ago

THAT the French should know better than any other nation how to enlist art in the service of science is just what might be expected. Such a service on the part of art to science is a fair return for the immense resources which scientific research has been able to place at the disposal of art. Nowhere have the discoveries of science been more useful or more utilised than at the celebrated porcelain manufactory of Sèvres, and the illustrations which we give to day will afford some idea of the beautiful results which are thus produced. As a permanent record of successful scientific efforts, nothing could be more satisfactory and appropriate. In the figure the characteristic features of the Arctic regions are rendered with almost perfect success and truthfulness.

From *Nature* 26, 36, 11 May 1882.



Sèvres Vase, commemorative of the North-East Voyage of Baron Nordenskjöld.

1. Prusiner, S. B. *Science* **216**, 136 (1982).
2. Kimberlin, R. H. in *Comparative Diagnosis of Viral Diseases* Vol. 3A (eds Kurstak, E. & Kurstak, C.) 349 (Academic, New York, 1981).
3. Hunter, G. D. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 365 (Academic, New York, 1979).
4. Lattarjet, R. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 387 (Academic, New York, 1979).
5. Millson, G. C. & Manning, E. J. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 409 (Academic, New York, 1979).
6. Rohwer, R. G., Brown, P. W. & Gajdusek, D. C. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 456 (Academic, New York, 1979).
7. Griffiths, J. S. *Nature* **215**, 1043 (1967).
8. Dickinson, A. G. & Fraser, H. in *Slow Transmissible Diseases of the Nervous System* Vol. 1 (eds Prusiner, S. B. & Hadlow, W. J.) 367 (Academic, New York, 1979).
9. Kimberlin, R. H. *Nature* **278**, 303 (1979).
10. Bruce, M. E. & Dickinson, A. G. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 71 (Academic, New York, 1979).
11. Kimberlin, R. H. & Walker, C. A. *J. gen. Virol.* **37**, 487 (1978).
12. Diener, T. O. *Viroids and Viroid Diseases* (Wiley, New York, 1979).
13. Dickinson, A. G. & Outram, G. W. in *Slow Transmissible Diseases of the Nervous System* Vol. 2 (eds Prusiner, S. B. & Hadlow, W. J.) 13 (Academic, New York, 1979).
14. Stanley, W. M. *Science* **81**, 644 (1935).

REVIEW ARTICLE

Aspects of the electrochemistry of steel in concrete

C. L. Page

Department of Construction and Environmental Health, University of Aston, Gosta Green, Birmingham B4 7ET, UK

K. W. J. Treadaway

Building Research Establishment, Garston, Watford WD2 7JR, UK

Information available on mechanisms of corrosion of steel in concrete is reviewed and the need emphasized for a better understanding of the electrochemistry of the subject and aspects of the physical performance of the concrete cover in relation to the diffusion of both chloride and oxygen. The practical influence of data obtained from studies in such areas in relation to longer-term improvements in durability of reinforced concrete is considered.

THE practice of embedding steel bars in a mixture of Portland cement, graded aggregates and water to form reinforced concrete was introduced during the second half of the nineteenth century. Initially the material was employed for projects of modest scale but the scope of its applications grew rapidly as its composite properties were appreciated.

Since the turn of the present century reinforced concrete has been in widespread use in a great variety of building and civil engineering work. It is exposed to many different types of environment and, as for any structural materials, the related characteristics of durability and need for maintenance are of considerable importance as service lives are expected to extend over several decades. In these respects, the performance of reinforced concrete has generally been impressive and there are many examples of its long-term durability (see ref. 1).

There are, however, several degradative processes which affect a minority of reinforced concrete structures leading to loss of serviceability or, in extreme cases, to structural collapse and, amongst these, the most common cause of deterioration is corrosion of reinforcing steel². This gives rise to a familiar sequence of events owing to the volume increase associated with the transformation of steel to rust which generates tensile stress within the surrounding concrete and causes cracking and eventual spalling of the cover to the reinforcement. Some examples of the types of problems and potential safety hazards which reinforcement corrosion can cause in buildings are illustrated in Fig. 1. Similar corrosion problems are also of major importance worldwide in relation to reinforced concrete road bridges and marine structures.

Despite the length of time during which experience has been gained of the corrosion behaviour of steel in concrete, there are still several important aspects that are not fully understood: in particular, the interpretation of information obtained from relatively short-term electrochemical studies, which are frequently used as an aid in predicting long-term performance³. Because such investigations often represent the only practicable approach for rapid assessment of the effects on corrosion behaviour of changes in materials specifications, concreting practices and so on, it is important to seek improved fundamental understanding of the electrochemistry of this complex and variable system.

We now survey briefly the commonly encountered causes of reinforcement corrosion and highlight some main areas where basic research is considered necessary.

Mechanisms of corrosion protection in concrete

Any attempt to describe in general terms the nature of concrete, viewed as a potentially corrosive medium, is likely to prove inadequate because many factors can affect both the physical pore structure of the material and the chemical composition of the solution phase within the pores. The most important constituent is the hydraulic binder, Portland cement, which, in its unhydrated form, consists mainly of four minerals of the following approximate compositions: tricalcium silicate (3CaO SiO_2), dicalcium silicate (2CaO SiO_2), tricalcium aluminate ($3\text{CaO Al}_2\text{O}_3$), tetracalcium aluminoferrite ($4\text{CaO Al}_2\text{O}_3 \text{Fe}_2\text{O}_3$). (In cement chemistry these minerals are conventionally referred to as C_3S , C_2S , C_3A and C_4AF .) When Portland cement is mixed with water, the constituent minerals undergo a complex sequence of hydration reactions, which gradually transform the paste to a hardened matrix of hydrated products. The most important reactions, insofar as the development of physical structure and mechanical properties are concerned, are those involving C_3S and C_2S which are converted to a gel of calcium silicate hydrates of ill-defined composition and structure (C-S-H gel) and to calcium hydroxide produced mainly in the form of well-developed portlandite crystals⁴. In a mature (fully hydrated) sample, C-S-H gel would be expected to account for ~70% of the weight of solid material and calcium hydroxide for ~20% (ref. 5), the eventual porosity and pore size distribution being determined largely by the water/cement ratio, as illustrated in Fig. 2.

From the point of view of the electrochemical behaviour of steel in concrete, the most significant feature of cement hydration is that the aqueous phase rapidly acquires a high pH value. Initially, the solution in contact with the hydrating cement grains contains hydroxides and sulphates of calcium, sodium and potassium but sulphate ions are rapidly precipitated to form the highly insoluble calcium sulphonyl-aluminate hydrate ($\text{C}_3\text{A} \cdot 3\text{CaSO}_4 \cdot 31\text{H}_2\text{O}$) (ref. 6). Thereafter, the liquid becomes increasingly concentrated with respect to sodium and potassium hydroxides and, as the pH rises, the concentration of calcium ions decreases in approximate conformity with the solubility product of calcium hydroxide. This state of affairs is reflected in the fact that, after a few weeks of hydration, pH values well in excess of 13 are frequently recorded for Portland cement pastes of water/cement ratios such as are commonly used in practice⁷⁻⁹.



Fig. 1 *a*, Cracking and spalling in the soffit of precast concrete roof planks resulting from reinforcement corrosion and leading to a loss of serviceability requiring roof replacement. *b*, Corrosion revealed when cracked concrete cover was removed from a number of precast concrete fins in a multi-storey car park. *c*, Propping of a staircase necessitated by structural weakening as a result of reinforcement corrosion (in a Middle Eastern structure). *d*, Spalling of cover to reinforcement in beams, and cracking of columns, as a result of reinforcement corrosion in a Middle Eastern building.

For iron in an alkaline environment, the Pourbaix diagram indicates that the metal may remain passive over a wide range of potentials¹⁰ and a very limited supply of oxygen is sufficient to provide the necessary degree of anodic polarization to maintain conditions within this range, as shown by rest potential measurement in deaerated limewater¹¹. Thus, in all but highly anaerobic environments, the primary mechanism of corrosion protection in concrete involves passivation of embedded steel resulting from the chemically inhibitive influence of hydroxyl ions.

The secondary feature of corrosion protection that is provided by dense concrete depends on the capacity of the material to act as a physical barrier between embedded reinforcement and the external environment, limiting the access of substances which may play a part in corrosion. In this respect, however, it is clear that concrete, with its continuous pore system and tendency to form surface cracks, is a far from perfect barrier. Thus concrete is not generally effective in excluding water but this may not cause corrosion problems unless the *pH* falls below that required to maintain passivity. Where the *pH* does fall, however, corrosion problems can arise even when the material is exposed to the air inside buildings provided the appropriate circumstances prevail¹². It has also been established that dense concrete is normally fairly permeable to oxygen and, although the rate of oxygen diffusion is obviously limited by increase in the degree of saturation of the material, a significant flux can be sustained even through specimens that are fully submerged for long periods in water¹³. The real importance of the physical

barrier presented by concrete cover is therefore related mainly to its ability to preserve the conditions of high *pH* needed to maintain the reinforcement in a passive condition by limiting the rate of ingress of acidic substances from the external environment.

Generally the acidic substances concerned are atmospheric carbon dioxide and, in polluted locations, other gases such as sulphur dioxide. These react with the alkaline constituents of the cement paste to form a carbonated zone which gradually penetrates into exposed concrete, reducing the *pH* of the affected region to a value below *pH*9.

Loss of protection by carbonation

The kinetics of carbonation for various types of concrete in different environments have been widely investigated and it has generally been observed¹⁴ that the relationship between carbonation layer thickness (*x*) and exposure time (*t*) is approximately parabolic in form:

$$\frac{dx}{dt} = kt^{-1/2}$$

The magnitude of the proportionality constant (*k*) depends on several variables related to the quality of the concrete (water/cement ratio, type of cement and so on) and to the environment (temperature, relative humidity and so on). By specifying a suitable grade of concrete for a given service situation, it is possible to ensure that the rate of advance of

carbonation dx/dt declines within a fairly short time to a low level ($<1 \text{ mm yr}^{-1}$). Provided the depth of concrete cover to the reinforcement is adequate, therefore, carbonation should not penetrate to an extent which endangers the passivity of the steel during the design life of a structure. It is this principle that underlies recommendations made in codes of practice regarding minimum depths of cover appropriate for concretes of different quality exposed to environments of varying severity¹⁵.

It is particularly important when concrete is being produced in adverse climatic conditions that adequate precautions be taken to ensure proper curing of the external surfaces because the formation of a weak, permeable surface layer would otherwise tend to promote carbonation¹⁶. In general, however, problems of corrosion arising from carbonation should be avoidable in normal density structural concrete, provided that well-established guidelines for constructional practice are followed. The situation that most frequently gives rise to corrosion risks, altogether more serious and difficult to contend with, occurs when reinforced concrete contains a substantial amount of chloride ion as this may cause depassivation of steel even in a medium of high pH such as uncarbonated concrete.

Loss of protection by chloride

Chlorides may be introduced into concrete during manufacture or service. Within the former category, there is the possibility of deliberate inclusion of admixtures containing calcium chloride to accelerate the early stages of cement hydration, a practice that is now strongly discouraged for reinforced concrete¹⁷. There are also many cases of adventitious introduction of chlorides as contaminants of the aggregates or mixing water, which may be almost unavoidable in certain circumstances where the use of locally available resources is necessary^{18,19}. During service, chlorides from external sources may penetrate concrete to levels which exceed normal depths of cover to reinforcement within a small fraction of the design life in the case of structures exposed to marine conditions^{20,21} or to de-icing salts^{22,23}.

Several matters arising in relation to the behaviour of steel in chloride-bearing concretes remain incompletely resolved despite long and numerous researches; the corrosive effects of calcium chloride in concrete were, in fact, studied as long ago as 1923 (ref. 24). We now consider some of the outstanding difficulties.

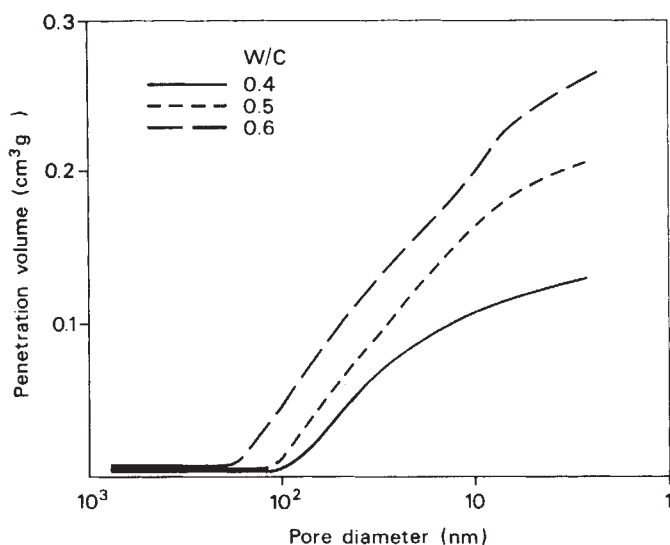


Fig. 2 Pore size distribution curves for mature hydrated ordinary Portland cement pastes determined by mercury intrusion porosimetry (Penetration volume indicates cumulative pore volume intruded by mercury; W/C indicates water: cement ratio of pastes).

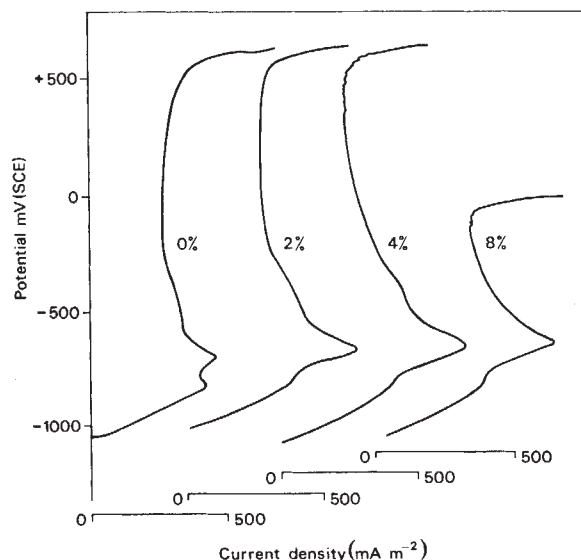


Fig. 3 Potentiodynamic anodic polarization of mild steel in hydrated ordinary Portland cement pastes of water/cement ratio 0.4 containing various percentages of calcium chloride dihydrate by weight of cement (sweep rate = 1 mV s^{-1}).

A question of practical importance, that has to be considered when chloride-bearing concreting materials are specified or when investigations of the durability of reinforced concrete structures are made, concerns the threshold level of chloride ion which is liable to stimulate corrosion of steel in concrete. As might reasonably be expected, however, bearing in mind the range of possible variations in concrete quality and exposure conditions, there is no simple answer to this question and on the basis of surveys of large numbers of reinforced concrete structures in Britain, the Building Research Establishment has recently proposed a classification for assessing the risk of corrosion²⁵. Thus for concrete made from ordinary Portland cement (or similar), it is considered that a low risk is associated with chloride contents (by weight of the cement) of $<0.4\%$, an intermediate risk for chloride contents of $0.4\text{--}1.0\%$ and a high risk for levels $>1.0\%$. The influence of carbonation in this process is also considered, the risk being enhanced when carbonation reaches the reinforcement.

Whilst providing sound practical guidelines that are adequate for many purposes, the above classification is not based on any mechanistic interpretations of the factors which determine the risk of corrosion in concrete containing a given quantity of chloride ion. It is therefore limiting in the sense that it provides no means for distinguishing *a priori* between the effects of a number of possible sources of variation, for example, the following: (1) chlorides associated with different cations; (2) chlorides introduced at or after the time of mixing; (3) chlorides introduced in the presence of other contaminant anions such as sulphates; (4) chlorides in concrete made from cements of different compositions; (5) chlorides in concrete of varying water/cement ratio; (6) chlorides in concrete cured or exposed in different conditions.

Because these and other factors are likely to influence the corrosion risk associated with the presence of a given concentration of chloride ions, there is a need to examine their roles and relative importance.

Mechanisms of corrosion of steel in concrete

The simplest and most reliable method involves long-term exposure tests on laboratory-prepared specimens of reinforced concrete, a costly and time-consuming procedure but which yields results that can generally be considered as a realistic guide to practical performance. The Building Research Establishment have used this approach in studying the influence of

many factors, such as the use of various types of reinforcing steel, and found that differences in performance may take several years to manifest themselves clearly³.

The alternative methods, that are often used when information has to be obtained within a restricted time, involve the application of various electrochemical techniques. They may be applied to steel which is either embedded in concrete, mortar, cement paste and so on, or else immersed in a solution of which the composition is believed to resemble that of the pore electrolyte in the concrete under consideration. With all of these methods, there are considerable difficulties in relating the results obtained to the performance of reinforced concrete as recorded from service experience or from exposure tests.

In studies involving the use of simulated pore-solutions there are fairly obvious uncertainties regarding the nature and concentrations of solutes that should be present. Thus it has commonly been assumed that saturated calcium hydroxide solution, which has a pH value of ~ 12.6 at 25°C , may be taken as an approximate electrolyte for electrochemical studies. This, however, is apparently at variance with the findings of several investigations of the composition of pore solutions extracted from actual Portland cement pastes or mortars; these have shown that, after a few weeks of curing, the pore liquid has often become essentially a mixed solution of sodium hydroxide and potassium hydroxide with a pH in the range 13–14 (refs 7–9). The latter electrolyte would be expected to maintain the passivity of steel in the presence of chloride concentrations well in excess of the level (~ 500 – 700 p.p.m.) which induces depassivation in saturated calcium hydroxide^{11,26}.

Before concluding that a more alkaline solution would constitute a more realistic model, however, note that the pore-solution in concrete is not buffered at a pH value in excess of $\sim \text{pH } 12.6$ and several commonplace processes may tend to reduce the naturally high alkalinities produced by cements rich in sodium and potassium oxides. Examples of such processes are: (1) use of calcium salts, for example, calcium chloride, as admixtures which would lead to precipitation of hydroxyl ions to form additional calcium hydroxide; (2) ingress of cations which form insoluble hydroxides from certain environments, for example magnesium ion from seawater; (3) leaching of sodium and potassium hydroxide from concrete frequently or continuously exposed to water; (4) diffusion of hydroxyl ion from the interior towards the carbonated surface zone of concrete exposed to the atmosphere; (5) simple dilution of the pore-solution associated with the cement paste by water from continuous pores within the aggregates.

It is perhaps reasonable, therefore, to consider that, in service conditions, the pH of uncarbonated concrete at the level of reinforcing bars may often tend towards the buffering range of saturated calcium hydroxide²⁷. This suggestion would fit in with the otherwise surprising observation that cements with high alkali-contents have not been found to provide significantly enhanced corrosion protection to steel in reinforced concrete exposed for long periods to chloride-bearing environments²⁸.

Apart from the question of pH, the other principal uncertainty regarding the composition of the liquid phase in concrete concerns the extent of binding of the chloride ions by solid hydration products. It is well-known that the aluminate and ferrite constituents of Portland cements can form incongruently soluble calcium chloro-aluminate and chloro-ferrite hydrates the former being the more important from the point of view of their capacity to remove free chloride ions from solution^{29,30}. There is also considerable evidence that the C_3A content of the cement has an important role in determining the extent to which chlorides included in the mix constituents are complexed and that this influences the degree of protection afforded to embedded steel in the presence of chloride^{29,31,32}; this is recognized in codes of practice by the imposition of stringent controls over the use of chloride-containing materials in concretes manufactured from cements of specified low C_3A content, such as sulphate resisting Portland cement¹⁷.

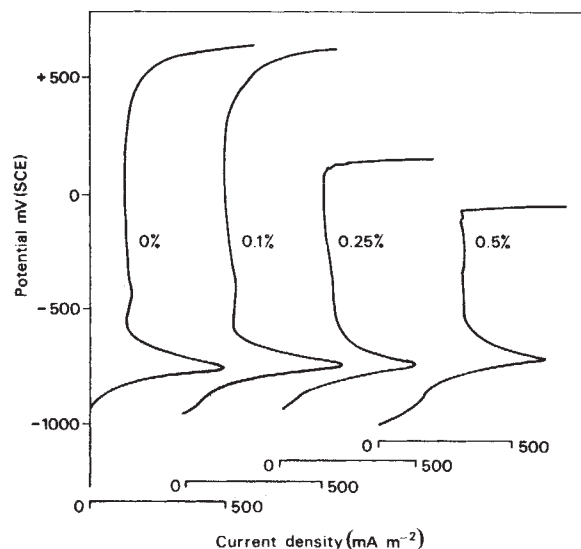


Fig. 4 Potentiodynamic anodic polarization of mild steel in saturated calcium hydroxide containing various percentages of calcium chloride dihydrate (sweep rate = 1 mV s^{-1}).

Nevertheless, there is a general lack of information about the levels of chloride likely to remain uncombined for long periods in the solution phase associated with concretes made from various cements in different conditions of natural exposure. The source of the chloride ions may be significant and it has been suggested that chlorides which penetrate concrete from the external environment tend to remain uncomplexed to a greater degree than those present at the time of manufacture^{21,33}. It is also unclear to what extent the presence of various levels of sulphate ions, either as constituents of the mix materials or as contaminants entering from the service environment, may liberate chloride ions owing to preferential formation of calcium sulpho-aluminate hydrates. Similarly the extent to which carbonation may cause decomposition of calcium chloro-aluminate hydrates is uncertain.

The above considerations highlight the need for further investigations of the solution chemistry of concrete pore electrolytes, aimed particularly at resolving the role of cement compositional variables (such as C_3A content and alkali content) in relation to several of the possible long-term changes that have been identified. Examination of solutions extracted from concretes, exposed for long periods to different service environments, is practicable and would appear to be an important field for future work.

Whilst improved understanding of the constitution of the aqueous phase within concrete would greatly facilitate attempts to understand the electrochemistry of steel in concrete by means of experiments involving steel electrodes immersed in simulated pore-solutions, the latter can never be an entirely satisfactory model. The main objection is that the kinetics of diffusion of the reactants and products of electrode processes cannot be comparable in solutions and in concrete.

As has been noted previously^{16,27}, depassivation as a result of chloride-induced pitting of steel in concrete is likely to be controlled by the rates of diffusion of several substances within the material. One significant factor is the rate of oxygen transport to the metal surface as this will influence cathodic polarization and determine whether the potential of the steel is sufficiently noble to support the formation and growth of pits at a given effective concentration of chloride and pH (ref. 34). For reinforced concrete, fully submerged in a chloride-bearing aqueous environment, pitting is often likely to be suppressed²⁷ and the only question that arises is whether the supply of oxygen to the steel will be even sufficient to balance the very low anodic leakage current density required to maintain the passive film³⁵ or whether a low-potential 'active' condition will result^{36,37}. It

has also been suggested that the magnitude of this leakage current density is itself controlled by diffusional constraints on the passage of iron ions into the surrounding medium, and may vary in response to change in the surface concentrations of chloride and hydroxyl ions³⁸. In conditions other than those of total immersion, the supply of oxygen to embedded steel is unlikely to be a limiting factor and the availability and rates of replenishment of chloride and hydroxyl ions seem liable to govern the anodic behaviour of the metal. Notwithstanding the controversy regarding mechanisms of chloride-induced pit initiation, it is generally accepted that pit growth is accompanied by the development of solution conditions within the pits of locally elevated chloride ion concentration and depressed pH (ref. 34). Factors which limit the diffusion of chloride ions or enhance the diffusion of hydroxyl ions in the vicinity of incipient pits will therefore tend to promote repassivation. One such factor is associated with the composition and physical barrier-layer characteristics of the lime-rich interfacial zone of segregation which surrounds steel embedded in cement-rich dense concrete^{16,27}; others are related to the mineralogy and pore structure of the hydrated cement matrix and, in this regard, it is significant that effective diffusivities of chloride ion, measured at fixed temperature for different cement pastes of constant water/cement ratio, have been found to vary widely^{39,40}. Activation energies recorded for chloride ion diffusion in hydrated cement pastes indicate that the kinetics are governed by interactions between the diffusing ions and the pore surfaces^{39,40}. The nature of these interactions and the effects on them of cement compositional variables represent a potentially useful field of continuing research. This may have practical applications related to the development of hydraulic cements which offer a high degree of corrosion protection to embedded steel by virtue of their capacity to retard the diffusion of chloride ions^{41,42}.

The foregoing discussion has indicated several major shortcomings of electrochemical investigations performed with electrodes immersed in solutions. To simulate the various dynamic factors which may play a part in controlling the durability of reinforcement in concrete, it is therefore necessary to undertake studies with electrodes embedded in concrete, mortar or cement pastes and several attempts have been made to carry out electrochemical measurements with specimens of these types. As will be discussed, the results obtained must generally be interpreted with considerable caution.

One of the greatest practical problems associated with all such investigations concerns the method of masking the embedded electrode so that it may be connected to external apparatus with a predetermined part of its surface area exposed to the electrolytic medium. The avoidance of crevice effects in the specimens is very difficult and the situation is exacerbated by the fact that many commonly used lacquers, coatings and gasket materials are themselves likely to suffer some degree of degradation during exposure for long periods to a highly alkaline environment such as concrete. When quantitative data are sought from long-term electrochemical studies of this nature, specimen design is therefore crucial if experimental artefacts are to be avoided. In particular, we suggest that, in planning such experiments it is worthwhile to consider using a range of specimens with systematically varied exposed surface area as this may facilitate the subsequent diagnosis of misleading effects of the type mentioned.

In the case of standard methods, such as potentiodynamic or potentiostatic polarization curve determinations, the data obtained will depend on variables such as the rate of change of potential and they have little prospect of yielding information which can be directly related to long-term durability. Their value is purely comparative in the sense that they can be used to provide rapid evidence of trends associated with a particular variable which is altered whilst others are held constant. Thus, for example, using fixed experimental conditions, different types of alloy steel may be graded according to their relative susceptibilities towards pitting in particular chloride-containing mortars³, or the relative effects of given dosages of chloride may

be examined for steel embedded in various types of cement pastes and mortars⁴³⁻⁴⁵.

As the typical data shown in Fig. 3 illustrate, however, the levels of chloride that are observed to cause significant effects in short-term experiments of this type are often considerably greater than those which are associated with a high risk of corrosion in long-term exposure tests or service. Note that the concentration of chloride ions in the original mixing water of the paste specimen containing 4% of calcium chloride dihydrate by weight of the cement, which appears to be on the borderline between passivity and pitting (see Fig. 3), is almost 1,000 times larger than the concentration required to induce pitting in saturated calcium hydroxide solution (see Fig. 4). This scale of disparity cannot be adequately explained in terms of the effects of chloro-aluminate hydrate formation or the excess hydroxyl ions contributed by sodium and potassium oxides in the cement, particularly bearing in mind the pH-depressing action of calcium chloride⁴⁴. The main factor is probably the slowness of chloride diffusion towards the embedded steel surfaces, which causes pitting to be 'missed' even at a relatively low rate of potential scan unless very large chloride concentrations are present in the pore solution.

Similar considerations apply, in general, to the use of galvanostatic polarization for examining the response of embedded steel electrodes. The technique has been useful as a means of providing comparative data^{3,46-50}. There is little doubt, however, that whilst this approach may be more reliable than potentiostatic or potentiodynamic methods when small current densities are applied, concentration polarization distorts anodic behaviour to an extent which makes accurate prediction of long-term performance impossible. A further disadvantage, common to galvanostatic, potentiodynamic and potentiostatic polarization curve determinations, is that the specimens used may be examined once only because of the likelihood of irreversible changes being brought about by potential attenuation over a range of several hundred millivolts; this effectively eliminates the possibility of following time-dependent changes in the response of individual specimens.

There are various alternative electrochemical methods which have been applied to steel in concrete or similar media in attempts to provide data continuously or at timed intervals over a period of several months or more. The range of techniques used has included rest potential measurements^{50,51}, current measurement in potentiostatic conditions in deaerated media³⁵, polarization resistance determination^{52,53}, and a.c. impedance measurement^{54,55}. In addition to the practical difficulties mentioned regarding electrode design to avoid misleading crevice effects and so on there are other grounds for caution in interpreting data from long-term electrochemical investigations carried out in concrete and, even in straightforward rest-potential studies, ambiguities can exist as other workers have

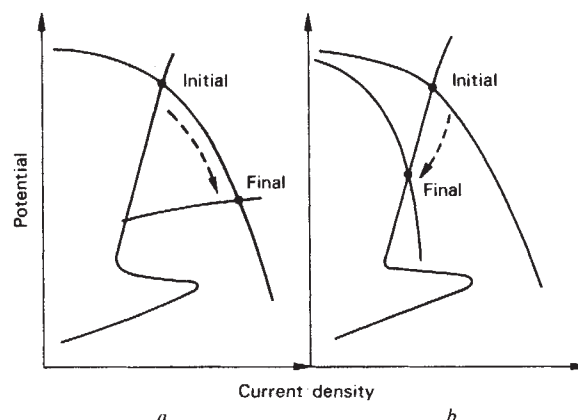


Fig. 5 Movements in rest-potential towards more negative values associated with: a, passive film breakdown; b, restriction of oxygen access.

indicated^{36,50}. Thus, it is well-established that passivity is frequently associated with a noble potential, which tends to increase with time towards a steady value, usually in excess of about -220 mV SCE (ref. 51) as the passive film thickens, implying a plentiful supply of oxygen. In cases where the rest-potential is observed to decrease with time, depassivation may or may not be implied since the increased cathodic polarization may be associated either with passive film breakdown (Fig. 5a) or with progressive restriction in the supply of oxygen to the surface of passive steel (Fig. 5b).

Of the other techniques referred to, it would appear that a.c. impedance studies offer the greatest likelihood of resolving present uncertainties about the fundamental nature of mechanisms of corrosion rate control in chloride-bearing concretes and similar media. Their major advantage in this respect lies in the ability to yield information concerning dynamic characteristics of the system, revealing in particular the influence of diffusion-controlled processes on the response of an embedded electrode. In view of the importance which has been ascribed to diffusional effects with respect to the mechanism of chloride-induced pit growth, it would be of considerable interest to compare a.c. impedance data obtained in various concrete or hardened cement paste systems, for which chloride diffusion kinetics have been studied directly.

Implications for the practical situation

We have, so far, concentrated on developments of a better basic understanding of the mechanisms of corrosion of steel in concrete. However, there is also an important practical need for such better understanding. The minority of reinforced concrete structures which deteriorate prematurely as a consequence of reinforcement corrosion do so as a result of poor constructional practices, inadequate design and, on occasions, poor quality of materials. While codes of practice lay considerable emphasis on design and general aspects of quality control, in respect to the use of marginal quality concrete materials they give little detailed guidance; it is here therefore where a better knowledge of corrosion protection mechanisms especially in relation to the microstructure of the matrix, interactions of contaminants (particularly chloride and sulphate) and the role of water/cement ratio and curing could well lead to wider usage of poorer quality indigenous materials without adverse effects on durability. At the same time, building owners will continue to repair existing deteriorated reinforced concrete structures and better understanding of the performance of repairs is urgently needed⁵⁶. This is particularly the case in relation to three factors: (1) The

diffusion of chloride through concrete (both in marine structures and where contact between chloride contaminated and uncontaminated units occurs). (2) The redistribution of anodic and cathodic sites in the concrete following repair (particularly in the substrate concrete). (3) The electrochemical interaction between repair and substrate materials and its influence on corrosion at the interface.

Conclusions

Work should be undertaken in the following areas to improve practical understanding of mechanisms of corrosion and protection of steel in concrete.

- Continued examination of the pore liquid in concrete at various stages of hydration and in different environmental conditions of curing and exposure.
- Further investigations of the distribution of chloride (at a range of concentrations) between the solid and solution phases of hydrating cements and the relationship of this distribution to cement type.
- Fundamental electrochemical studies of the performance of steel electrodes in electrolytes simulating pore solutions and complementary electrochemical studies on steel/concrete systems.
- Studies of the diffusion rates of chloride through uncontaminated cement pastes and through those contaminated with chloride and/or sulphate.
- Studies of oxygen diffusion through cement pastes and concretes, particularly in underwater exposure.
- Development of improved electrochemical monitoring techniques for laboratory experimentation and *in situ* measurement of corrosion rates in structures, which accommodate difficulties arising from the high electrolytic resistance of concrete, and from crevice effects at electrode edges.
- Investigation of the mechanisms of macrocell development in chloride contaminated concrete, especially in relation to concrete repairs.

Improved understanding in these areas would have wide scope for application to the construction of durable reinforced concrete structures in adverse conditions and to the evolution of successful repair techniques. The economic benefits to be gained from solution of the problems outlined above are therefore substantial.

This review is based in part upon research carried out at the Building Research Establishment of the Department of the Environment and this paper is published by permission of the Director. C.L.P. gratefully acknowledges the financial support received from the SERC.

1. *The Durability of Reinforced Concrete Buildings*, National Building Studies Spec. Rep. No. 25 (HMSO, London, 1956).
2. Midgley, H. C., Figg, J. W. & McLean, M. J. *Concrete* **7**, 24–26 (1973).
3. Brown, B. L., Harrop, D. & Treadaway, K. W. J. *Building Research Establishment Current Paper* CP 45/78 (BRE, Garston, 1978).
4. Lea, F. M. *The Chemistry of Cement and Concrete* 3rd edn, 71 (Arnold, London, 1970).
5. Diamond, S. *Proc. Conf. on Hydraulic Cement Pastes: Their Structure and Properties*, Sheffield, 2–30 (Cement and Concrete Association, Slough, 1976).
6. Roberts, M. H. *Building Research Establishment Current Paper* CP 61/68 (BRE, Garston, 1968).
7. Peguin, P., Rubaud, M., Longuet, P. & Zelwer, A. *Cah. Cent. scient. tech. Batim.* No. 130, Cahier 1109 (1972).
8. Longuet, P., Burglen, L. & Zelwer, A. *Revue Matér. Constr. Trav. publ. Cim. Bétons*, **676**, 35–41 (1973).
9. Barneyback, R. S. Jr & Diamond, S. *Cem. Concr. Res.* **11**, 279–285 (1981).
10. Pourbaix, M. *Atlas d'Equilibres Electrochimiques*, 307 (Gauthier-Villars, Paris, 1963).
11. Hausmann, D. A. *Matér. Prot.* **6**, 19–23 (1967).
12. RILEM Technical Committee CRC Report, *Matér. Struct. Test. Res.* **9**(51), 187 (1976).
13. Gjorv, O. E., Vennesland, O. & El Busaidy, A. H. S. in *Proc. Corrosion/76*, Pap. No. 17 NACE, Houston, 1976).
14. Roberts, M. H. *Building Research Establishment Information Paper* IP 6/81 (BRE, Garston, 1981).
15. Treadaway, K. W. J. *Proc. Symp. on Corrosion of Steel Reinforcements in Concrete Construction*, London, 1–14 (Society of Chemical Industry, London, 1979).
16. Page, C. L. *Bull. Inst. Corros. Sci. Technol.* **77**, 2–7 (1979).
17. British Standards Institution Code of Practice CP 110:1972 Part 1, *The Structural Use of Concrete* (as amended May 1977).
18. Fookes, P. G. & Collis, L. *Concrete* **10**, 14–19 (1976).
19. Crookes, R. N. *Proc. Symp. Corrosion of Steel Reinforcements in Concrete Construction*, London 31–42 (Society of Chemical Industry, London, 1979).
20. Gjorv, O. E. & Vennesland, O. *Cem. Concr. Res.* **9**, 229–238 (1979).
21. Browne, R. D. & Geoghegan, M. P. *Proc. Symp. Corrosion of Steel Reinforcements in Concrete Construction*, London, 79–104 (Society of Chemical Industry, London, 1979).
22. Collepardi, M., Marcialis, A. & Turriziani, R. *Il Cemento* **69**, 143–150 (1972).
23. Ost, B. & Monfore, G. E. J. *Res. Dev. Labs Portld Cem. Ass.* **8**, 46–52 (1966).
24. Cottringer, P. & Kendall, N. C. *Concrete* **22**, 150–155 (1923).
25. Everett, L. H. & Treadaway, K. W. J. *Building Research Establishment Information Paper* IP 12/80 (BRE, Garston, 1980).
26. Gouda, V. K. *Br. Corros. J.* **5**, 198–203 (1970).
27. Page, C. L. *Nature* **258**, 514–515 (1975).
28. Lea, F. M. & Watkins, C. M. *National Building Studies Research Paper* No. 30, (HMSO, London, 1960).
29. Roberts, M. H. *Mag. Concr. Res.* **14**, 143–154 (1962).
30. Richartz, W. *Zem.-Kalk-Gips* **22**, 447–450 (1969).
31. Verbeck, G. J. *Am. Concr. Inst. Publ. SP* 49, 21–38 (1975).
32. *Building Research Establishment The Structural Condition of Intergrid Buildings of Prestressed Concrete* (HMSO, London, 1978).
33. Vassie, P. R. W. *Bull. Inst. Corros. Sci. Technol.* **81**, 11 (1980).
34. Henriksen, J. F. *Corros. Sci. Technol.* **81**, 1241–1249 (1980).
35. Gronvold, F. O., Preece, C. M. & Arup, H. *Proc. 8th Int. Congr. Metallic Corrosion*, Mainz, 2, 1800–1804 (1981); *Steel in Concrete Newslett.* No. 3, 3–4 (1979).
36. Wilkins, N. J. M. & Lawrence, P. F. *Proc. Symp. Corrosion of Steel Reinforcements in Concrete Construction*, London, 105–118. (Society of Chemical Industry, London, 1979).
37. Wilkins, N. J. M. *Bull. Inst. Corros. Sci. Technol.* **81**, 8 (1980).
38. Gronvold, F. O. *Steel in Concrete Newslett.* No. 4, 4–5, (1979).
39. Collepardi, M., Marcialis, A. & Turriziani, R. *J. Am. ceram. Soc.* **55**, 534–535 (1972).
40. Page, C. L., Short, N. R. & El Tarras, A. *Cem. Concr. Res.* **11**, 395–405 (1981).
41. Short, N. R. & Page, C. L. *Proc. 8th Int. Congr. Metallic Corrosion*, Mainz, 2, 1767–1772 (1981).
42. Arup, H. *Steel in Concrete Newslett.* No. 7, 3–4 (1981).
43. Mitheux, I. P. thesis, New South Wales Institute of Technology (1976).
44. Page, C. L. *Paper presented at int. Seminar on Electrochemistry and Corrosion of Steel in Concrete*, Copenhagen (1979).
45. Baumel, A. & Engel, H.-J. *Arch. EisenhütWes.* **30**, 417–428 (1959).
46. Kaesche, H. *Zem.-Kalk-Gips*, **12**, 289–305 (1959).
47. Lewis, D. A. *Proc. 1st Int. Congr. Metallic Corrosion*, London, 547–555 (1961).

48. Gouda, V. K. & Monfore, G. E. *J. Res. Dev. Labs Portld Cem. Ass.* **7**, (3), 24–31 (1965).
 49. Gouda, V. K. & Halaka, W. Y. *Br. Corros. J.* **5**, 204–208 (1970).
 50. Vassie, P. R. *Transport and Road Research Laboratory Rep. SR 397* (Crowthorne, TRRL, 1978).
 51. Stratfull, R. F. *High. Res. Rec.* **433**, 12–21 (1973).
 52. Gouda, V. K., Shater, M. A. & Mikhail, R. Sh. *Cem. Concr. Res.* **5**, 1–13 (1975).

53. Locke, C. E. & Siman, A. *ASTM Publ. STP 713*, 3–16 (1980).
 54. Dawson, J. L., Callow, L. M., Hladky, K. L. & Richardson, J. A. in *Proc. Corrosion/78* (Paper No. 125, NACE, Houston, 1978).
 55. King, R. A., Dawson, J. L. & Geary, D. *Proc. Symp. Corrosion of Steel Reinforcements in Concrete Construction*, London, 135–143 (Society of Chemical Industry, London, 1979).
 56. O'Brien, T. *Proc. Instn civ. Engrs*, Part 1, **68**, 399–408 (1980).

ARTICLES

Identification of resonance features within the rings of Saturn

J. B. Holberg*, W. T. Forrester* & Jack J. Lissauer†

* Earth and Space Sciences Institute, University of Southern California, Tucson Laboratories, Tucson, Arizona 85713, USA

† University of California, Departments of Astronomy and Mathematics, Berkeley, California 94720, USA

The Voyager 2 UV spectrometer observed a stellar occultation by the rings of Saturn, which located ring features with an accuracy of 12 km. A high-resolution (3 km) optical depth atlas of the rings shows at least nine features, including four density wave patterns, identified with satellite resonances. Analysis of these density wave patterns yields the first surface mass densities for the A ring and, together with our optical depth atlas, a total ring mass of 6.4×10^{-8} Saturn masses.

It has long been customary to link certain gaps and boundaries observed in the rings of Saturn to the locations of resonances with several of Saturn's satellites. Most notable in this regard has been the association of the outer edge of the B ring with the location of the Mimas 2:1 resonance. The vastly increased spatial resolution of the rings available from Voyagers 1 and 2 has made it possible to pursue this activity with much more confidence; however, the new clearer picture of the rings is complicated by two factors. First, the rings contain a vast number of radial features on all scales. Second, the newly discovered inner satellites have greatly increased the number of potential resonances with which one has to contend. Both of these factors have led to the suggestion of numerous possible associations^{1–4}. To resolve any ambiguities a further increase in spatial resolution, together with a reliable distance scale for the rings is necessary. Both of these criteria are met by the stellar occultation experiment performed by Voyager 2.

Some of the first Voyager 1 images of the rings were used by Collins *et al.*¹ to produce a list of observed features and possible corresponding resonance locations involving Titan, Iapetus, Hyperion, Mimas and the co-orbital satellites 1980S1 and 1980S3. The uncertainty in the absolute location of ring features, however, was ± 500 to $\pm 1,000$ km. Later Voyager 1 and 2 images showed wave-like patterns associated with several A ring features. These features were suggested as being near to areas where two or more resonances coincided². A more positive identification was provided by Cuzzi *et al.*⁵ who analysed Voyager 1 images of the Cassini division showing density waves due to the Iapetus 1:0 resonance. Further evidence of density waves associated with the 1980S1 2:1 resonance was provided by the Voyager photopolarimeter experiment⁶ which obtained stellar occultation data simultaneously with the UV spectrometer.

The ring features which we describe here are located near the inner Lindblad resonances with the external satellites of Saturn. Many theoretical consequences of such resonances, including tightly wound spiral density waves propagating radially outward from the resonance location, were first discussed by Goldreich and Tremaine⁷ (also see refs 5, 8).

In general these resonances are described by a ratio of two integers of the form $l/(m-1)$, which represents the approximate ratio of the mean orbital motion of a satellite to that of a ring particle. Resonances of importance in Saturn's rings can be classified into two types¹. Type I resonances, where $l = m$, are

usually the strongest and are independent of satellite orbital eccentricity, while type II resonances, where $l = m + 1$, are weaker, being proportional to the first power of satellite orbital eccentricity. A special type I resonance case occurs when $l = m = 1$. Here the rate of the apsidal advance of a ring particle's orbit equals the mean orbital motion of the satellite. We will refer to the three resonances of this variety present in the rings as apsidal resonances. An additional class of resonance considered here involves resonances in the rings due to the rotation of Saturn⁴. Such resonances depend on an azimuthal asymmetry in the gravitational figure of Saturn.

The locations of the resonances used here are those of Lissauer and Cuzzi⁸ who have computed the locations and strengths of expected satellite resonances using the zonal harmonics of Saturn's gravitational potential through J_6 . These locations differ slightly from those given by Collins *et al.*¹ and except for the apsidal resonances are believed to be accurate to within ± 10 km. Expected resonance locations are shown in Fig. 1 together with normal optical depths (τ) of the A, C and portions of the B ring observed by the UV spectrometer.

δ Sco occultation

During the Voyager 2 Saturn encounter the UV spectrometer⁹ observed an occultation of the bright B0.5 IV star δ Scorpii by the rings of Saturn. The occultation began as the star emerged from the atmosphere of Saturn at 23:45:07 UTC 25 August 1981 and continued for 130 min. During this time the apparent path of the star traversed the entire ring system from the D ring to beyond the F ring with a mean radial velocity of 10.3 km s^{-1} . Throughout the occultation the UV spectrometer obtained a complete 912–1,700 Å spectrum of δ Sco every 0.32 s, yielding a resolution of ~ 3.3 km in the rings. A preliminary discussion of the UV spectrometer ring occultation results is given elsewhere¹⁰. The data from which the normal optical depths of the rings were derived are the result of summing an entire UV spectrum to yield a single intensity measurement at an effective wavelength of 1,125 Å. These intensity data were then corrected for instrumental background and the effects of spacecraft limit cycle motion. Normal optical depths have been computed from the ratio of the attenuated to the unattenuated signal and corrected for the 28.71° angle between the line of sight to δ Sco and the plane of the rings. Details of the data reduction will be presented elsewhere.

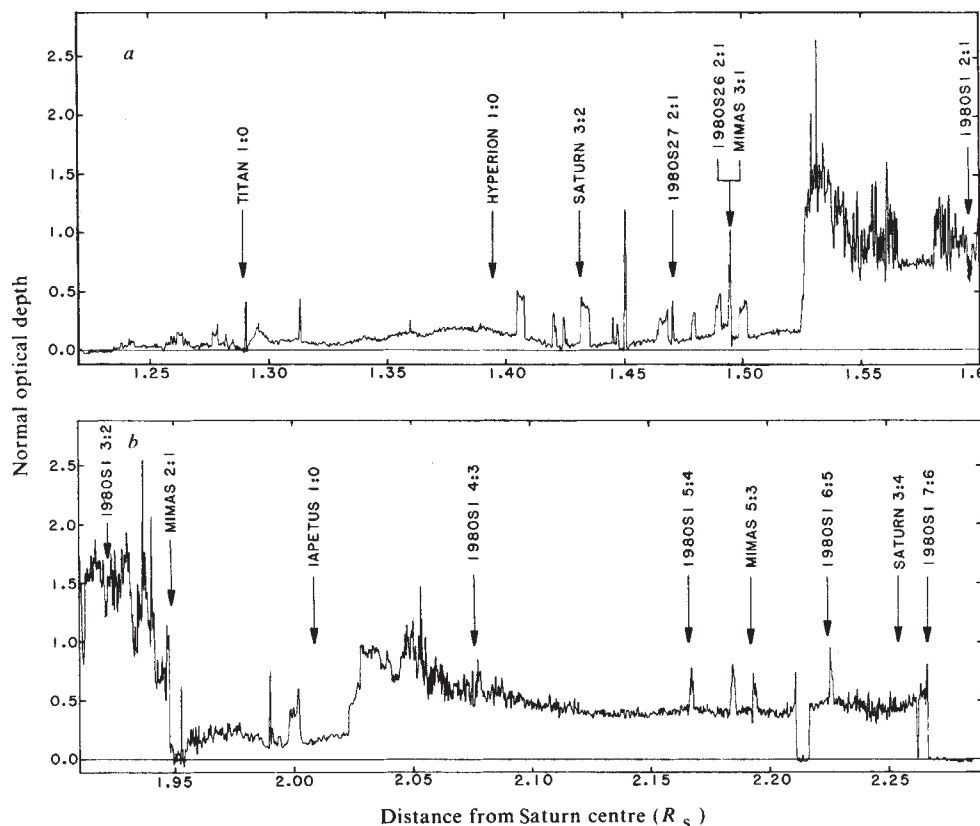


Fig. 1 The normal optical depths in the rings observed by the Voyager 2 UV spectrometer. These data, which have been plotted at a resolution of ~ 15 km, span the entire C and A rings and include those portions of the B ring containing resonances discussed in this report. The expected locations of these resonances^{4,8} are indicated by arrows. *a*, The C ring and the inner portion of the B ring. *b*, The outer portions of the B ring, Cassini division and the A ring. Optical depths between 1.95 and $1.983 R_S$ are preliminary due to present uncertainty in the spacecraft limit cycle in this region. The resulting uncertainty in the Cassini division mass (Table 2) is perhaps as large as 50%. Precise ring boundaries observed by the UV spectrometer are given in Table 2.

Many of the identifications of ring resonances presented here rest on the absolute radial locations obtained for the associated features. A brief discussion of the accuracy of these determinations is therefore appropriate. For a star occulted by the rings, the location of observed ring features depends only on the apparent position of the star, the precise trajectory of the spacecraft relative to Saturn coupled with the timing of events, and the location of Saturn's rotational pole. The uncertainty in the apparent position of δ Sco, after corrections for proper motion, parallax and aberration of starlight is negligible (<0.1 km in the ring plane). The trajectory uncertainty projected on to the ring plane is of the same order as the uncertainty due to the UV spectrometer time resolution, 2.4 and 3.2 km respectively. The dominant uncertainty, at present, is due to the location of Saturn's rotational pole. Errors in the location of ring features produced by this factor are difficult to estimate, but use of various pole locations derived from Voyager trajectory analysis as well as pre-Voyager determinations cause location changes of <20 km in the rings.

Fortunately, it is possible to compare directly the distance scale used here with the locations of features seen in two additional stellar ring occultations observed with the UV spectrometer from Voyagers 1 and 2. These are a Voyager 1 observation of ι Herculis and the Voyager 2 entrance of δ Sco through the opposite side of the rings. Both of these occultations extended out to $1.43 R_S$ ($1R_S = 60,330$ km). The locations of all recognizable features in these three occultations are in mutual agreement to within 12 km except for one ringlet known to be eccentric from imaging data. If we make the reasonable assumption that this agreement is not just coincidence but rather a consequence of symmetry in this part of the C ring, then we can be confident of our absolute distances to this level. Ultimately it should be possible to reduce the locational uncertainties to 6 km or better.

Detected and undetected resonances

The recently discovered small inner satellites of Saturn², 1980S1, 1980S3, 1980S26, 1980S27 and 1980S28 all possess numerous resonances within the ring system. Because of the small mass of several of these bodies only a fraction of these

resonances are important. Of these, the type I resonances due to the leading, more massive, co-orbital 1980S1 are strongly evident in the rings. Figure 2 shows full resolution data for the six type I 1980S1 resonances seen in Fig. 1. Three of these resonances display characteristic density wave structure (2:1, 5:4, 6:5) while the 3:2 and 4:3 resonances are more chaotic. These will be discussed more fully together with other resonances in high optical depth regions of the rings; however, one point is unique to 1980S1 resonances. The interchange of the orbits of 1980S1 and 1980S3 which occurs every 4 yr¹¹, changes the resonance locations of 1980S1 by 10–15 km. A 'kink' in the wave train of approximately this magnitude will develop (S. Tremaine, personal communication) and propagate outward at the wave's group velocity⁷. In the ~ 960 days (ref. 11) between the last interchange and our observations, these waves have travelled roughly 60 km from the 2:1 resonance of 1980S1 and over 100 km from the other type I resonances. This effect may cause a slight error in our subsequent determination of the surface density at the 2:1 resonance. The 7:6 resonance is coincident with the outer edge of the A ring. The smaller co-orbital, 1980S3, has resonances which occur 30–45 km inward of the corresponding 1980S1 resonances. These resonances are not apparent in our data.

Features due to resonances with the two F ring shepherd satellites 1980S26 and 1980S27 are also present in our data (Fig. 1). Many of the >50 type I shepherd satellite resonances in the rings are noticeable, particularly in the outer A ring. We have included in this discussion only the 2:1 resonances with each of these satellites because they occur in low optical depth regions. Because of their number and lower amplitude, discussions of the remaining 1980S26 and 1980S27 resonances will be presented elsewhere. There are ~ 100 1980S28 type I resonances present in the rings, near the outer edge of the A ring they reach a separation of 9 km. At present there is no evidence of any features due to resonances with this small satellite.

The occultation data show clear evidence (Fig. 3) of features corresponding to all three of the strong Mimas resonances present in the ring system. The strongest and best known is the 2:1 resonance which has long been believed to be coincident

with the outer edge of the B ring. The two additional Mimas resonances are the only known examples of type II resonances seen in the rings. Apparently, the 0.02 eccentricity of Mimas is large enough to allow Mimas to exert sufficient torque to create observable features of the 5:3 and 3:1 resonance locations.

The only apsidal resonance which we can definitely detect is the Titan 1:0 which occurs in the C ring (Fig. 4). Hyperion and Iapetus, being much less massive and more distant than Titan, exert significantly less torque at their apsidal resonances so that our failure to observe features at these resonances is not surprising. The density waves described by Cuzzi *et al.*⁵ near the Iapetus apsidal resonance are of too low an amplitude to be seen in our data. The stated location of the wave source⁵, however, is seen in Fig. 1 as the inner edge of the band of low τ containing the Iapetus resonance. The wave source and the resonance location are separated by ~ 370 km.

Finally, we have examined our data for the presence of ring features associated with the predicted location of resonances with Saturn's rotational period. The possibility of such resonances has been suggested by Franklin *et al.*⁴, who tentatively identified the locations of the 3:2 and the 3:4 rotational resonances with two prominent gaps in the rings seen in Voyager 1

images. Our data (Fig. 1) show the 3:2 resonance to be located at the outer edge, rather than the centre, of the gap suggested by Franklin *et al.*⁴. The 'gap' itself has a distinct non-zero optical depth and is similar in appearance to several other inter-ringlet regions in this part of the C ring. The precise location of the 3:2 resonance falls at the inner edge of the 230 km-wide ringlet located at $1.434 R_S$. Doubt as to the significance of any link between the resonance and either the gap or ringlet is increased by the failure to observe the presence of the related 3:4 resonance. Franklin *et al.*⁴ associated the 3:4 resonance with the prominent narrow gap between the Encke division and the outer edge of the A ring. We find this gap to be located at $2.2627 R_S$, 440 km outside the location of the resonance but within the ± 700 km uncertainty of the Franklin *et al.*⁴ observational data. At the location of the resonance itself we see no convincing evidence of any associated structure.

Resonances in low τ regions

The morphology of ring features associated with the strongest resonances appears to be controlled by the local optical depth

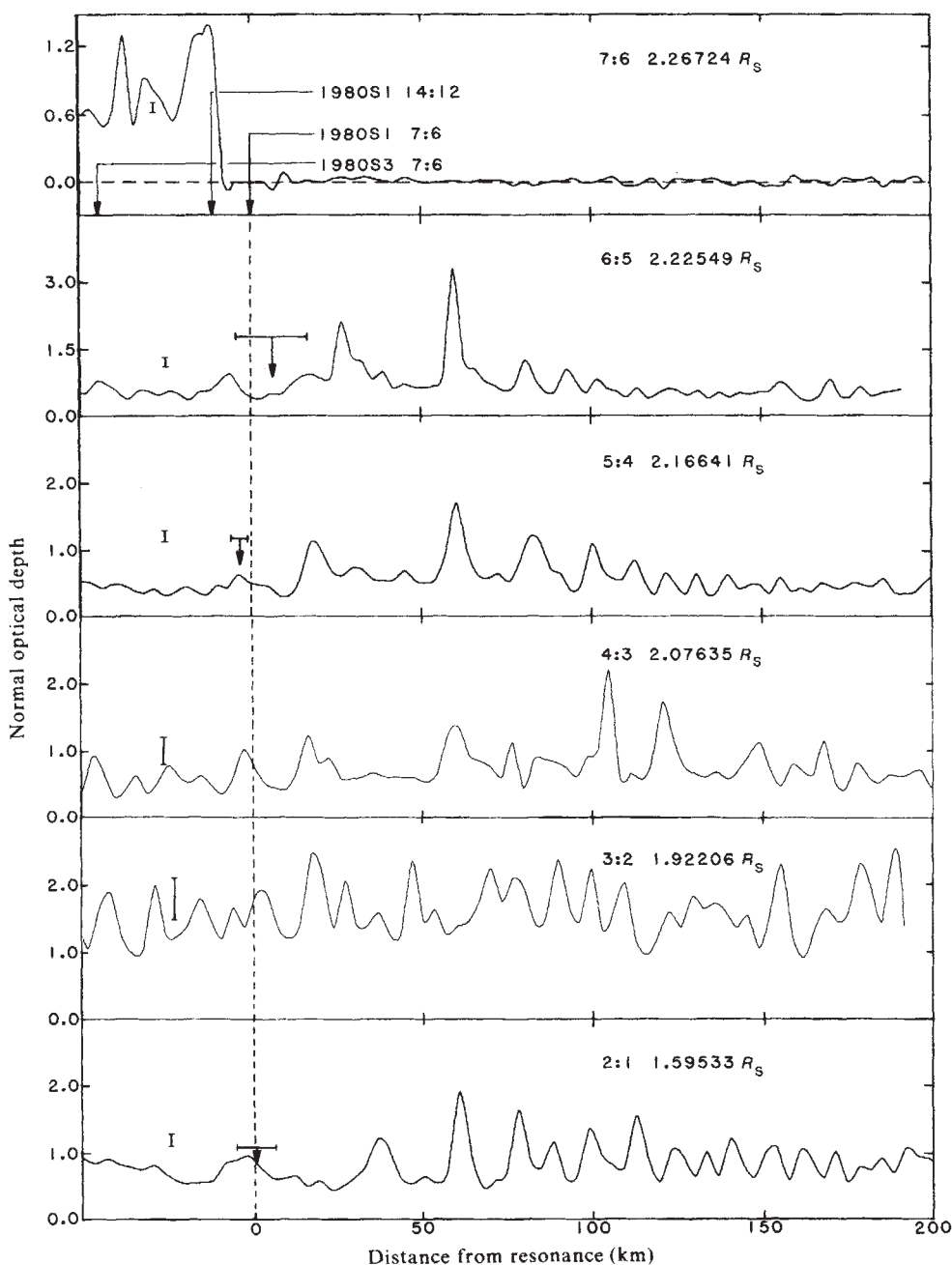


Fig. 2 The normal optical depths for regions located near the six type I resonances due to the co-orbital satellite 1980S1. The origin of the horizontal scale corresponds to the expected location of each resonance. Spiral density wave patterns with characteristic damped waves of decreasing wavelength are seen exterior to the 2:1, 5:4 and 6:5 resonances. The appearance of the rings at the 3:2 and 4:3 resonances is more chaotic. Vertical error bars show the single sample noise level at the indicated τ . Arrows and horizontal error bars show the resonance locations deduced from analysis of the density waves.

Table 1 Ring features corresponding to resonance locations

Source	Resonance	Expected location (R_s)*	Surface density σ (g cm^{-2})	Opacity K ($\text{cm}^2 \text{g}^{-1}$)	Type of feature
1980S27	2:1	1.47043			Narrow ringlet/gap
1980S26	2:1	1.49456			Narrow ringlet/gap
1980S1	2:1	1.59535	70 ± 10	0.01 ± 0.002	Density waves ⁶
	3:2	1.92206			?
	4:3	2.07635			Disturbance
	5:4	2.16641	56 ± 12 (100)	0.008 ± 0.002	Density waves
	6:5	2.22549	40 ± 6 (110)	0.01 ± 0.002	Density waves
Mimas	7:6	2.26724			Outer edge of A ring
	3:1	1.49503			Narrow ringlet/gap
	2:1	1.94847			Outer edge of B ring (narrow ringlet/gap)
	5:3	2.19286	45 ± 8 (80)	0.01 ± 0.002	Density waves
Titan	1:0	1.28951			Eccentric ringlet/gap
Hyperion	1:0	1.39499			None
Iapetus	1:0	2.00867			? (Density waves ⁵)
Saturn	3:2	1.43163			Outer edge of gap?
	3:4	2.25543			None

* $1R_s = 60,330$ km.

† Measured outside region of enhanced surface density, densities in parentheses refer to region of enhancement.

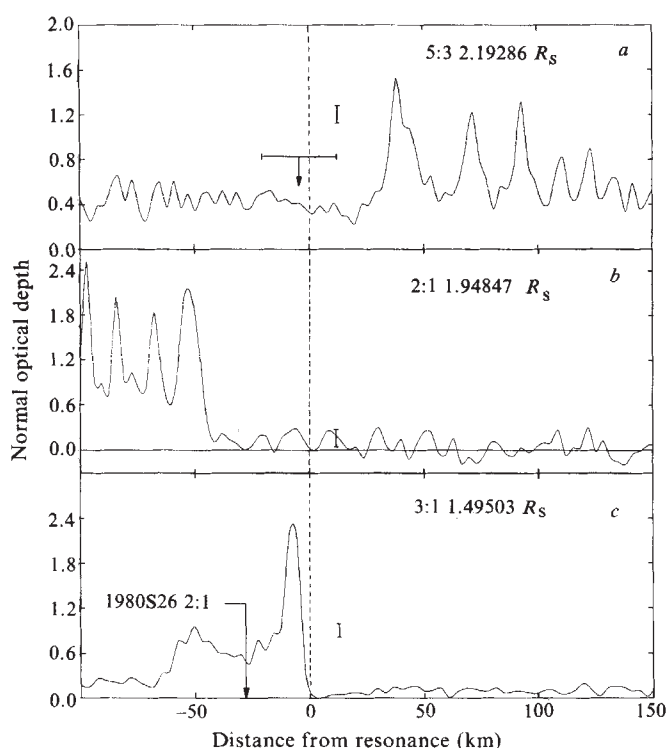


Fig. 3 The normal optical depths for regions located near the three Mimas resonances. The origin of the horizontal scale corresponds to the expected location of each resonance. The 5:3 resonance (a) is an excellent example of density waves propagating outwards from a resonance location. The 2:1 resonance (b) is located 48 km outside the outer boundary of the B ring but within the region over which this boundary is observed to vary³. The 3:1 resonance (c) is located at the outer edge of a narrow ringlet in the C ring and may interact with the 1980S26 2:1 resonance to produce the sharp peak in the ringlet. Note the near zero optical depth exterior to this ringlet. Vertical error bars show the single sample noise level at the indicated τ . The arrow and horizontal error bar near the 5:3 resonance indicate the location of the resonance deduced from analysis of the density waves.

of the region in which the resonances reside. Looking first at regions of low optical depth ($\tau < 0.2$), we note that there are three locations containing major resonances in the C ring (Fig. 1). The Titan apsidal resonance lies farthest inward, while the 2:1 resonances with 1980S27 and 1980S26 are located near the outer edge of the C ring, the latter of these two being very nearly coincident with the Mimas 3:1 resonance (see Fig. 3). All three of these resonances occur in regions characterized by a gap of lower than average optical depth containing a narrow optically thick ringlet. The gap associated with the Titan apsidal resonance is several hundred kilometres wide and completely empty. The 1980S27 resonance is situated in a region of lower than average, but apparently non-zero, optical depth while the region just exterior to the 1980S26 2:1 and Mimas 3:1 resonance is devoid of ring material to within our level of detection.

The Titan apsidal resonance, the only one of these features for which there exist multiple UV spectrometer observations, is associated with a narrow, double-peaked, optically thick ringlet of substantial eccentricity^{3,10}. The two Voyager 2 UV spectrometer δ Sco occultations locate this ringlet at 111 km and 67 km exterior to the Titan apsidal resonance at longitudes separated by 149° and 3.53 h of time. The 1980S27 2:1 resonance and the close pair formed by the 1980S26 and Mimas resonances are also associated with narrow, double-peaked, optically thick ringlets. The Mimas 2:1 resonance at the inner edge of the Cassini division also manifests itself in a gap containing an opaque double-peaked ringlet. In this case, not only is the ringlet eccentric but the inner edge of the gap also varies in radius and is shaped like an ellipse centred on Saturn³. In the UV spectrometer data the outer edge of the B ring was located 48 km inside the theoretical location of the Mimas 2:1 resonance. At the time of the stellar occultation Mimas, which controls the orientation of the B ring's outer edge, lagged the point of observation in the rings by a phase of 194° . This datum, when combined with imaging measurements, should give a more precise picture of the location and shape of the outer edge of the B ring.

We do not understand fully why resonances which occur in regions of low optical depth manifest themselves as gaps containing narrow double-peaked ringlets. Resonantly induced collisions are more likely to clear gaps in very low τ regions⁸, however, it is questionable whether τ is sufficiently low for this process to occur even in the C ring, also in most cases the size of the gap is larger than the collision predicted width. Density

waves may also be more effective in clearing a gap in low surface density regions because they will quickly become nonlinear, depositing their negative angular momentum over a smaller region⁷. A complete theory of these gaps must involve the opaque ringlets embedded within them. These ringlets, which seem to be characterized by a double-peaked structure similar to some of the uranian rings¹², also occur within some non-resonant gaps in the C ring. Imaging data will be important for the determination of the azimuthal structure of these ringlets as well as the gap edges.

Resonances in regions of medium and high τ

The predominant effect of resonances in regions where τ is between 0.4 and 1.0 is the excitation of density waves. We have selected for analysis the four clearest examples of such density waves; those associated with the Mimas 5:3 resonance (Fig. 3), and the 6:5, 5:4, and 2:1 resonances with 1980S1 (Fig. 2), the 2:1 having been previously studied by Lane *et al.*⁶. The two other 1980S1 type I resonances are noisier in our data. The 4:3 occurs in the inner and most opaque part of the A ring and clearly causes some stirring but a clean wave pattern is not present. The 3:2 occurs in a high τ region of the B ring and is not immediately evident in our data. At present it is unclear whether the lack of observable density waves in the highest τ regions is due to their absence, their confusion with the crowded ringlet structure of the B ring² or simply our low signal level when τ becomes very large.

One of the theoretical predictions concerning the development of density waves near a resonance is the opening of a gap just exterior to the resonance⁷. Not only do we find no evidence of density waves clearing gaps, the average τ in resonance regions is in general higher than that of surrounding regions. Density waves seem to damp and thereby deposit negative angular momentum over a region of several hundred kilometres, so it may be that material is brought inwards towards a resonance faster than it can be cleared from the region by damping of the first wave. An equilibrium may be reached when surface mass density (σ) is increased enough for viscous diffusion out of the region to create a balance. Another possibility is that collisions alter the particle size distribution by increasing the numbers of small particles, thereby increasing τ .

The density waves excited at the Mimas 5:3 and at the 1980S1 type I resonances are also quite highly nonlinear (shocked) for the first few wavelengths. These waves have a distinctly non-sinusoidal character with a sharp peak at the location of the shock and a broad flat trough (see ref. 13 for a theoretical analysis of nonlinear density waves in galaxies).

Density waves are a very useful tool for deducing such ring properties as surface mass density and viscosity, unmeasurable by any other currently available means^{5,6}, however, some caution must be exercised. First, the shocking of the waves removes a great deal of their angular momentum, thus viscous damping plays a minor part and any measure of viscosity from wave damping would only constitute a crude upper bound (a slight phase shift also results when the waves shock, thus changing the apparent wavelength between crests, however, this small effect should not significantly affect our estimates of σ derived in equation (1)). Second, the mean (wavelength-averaged) surface density can be non-uniform over the region of a particular wave train (we see evidence of this in the A ring). This may lead to problems in the determination of the resonance location as the mean resonance wavelength decreases as σ/d , where d

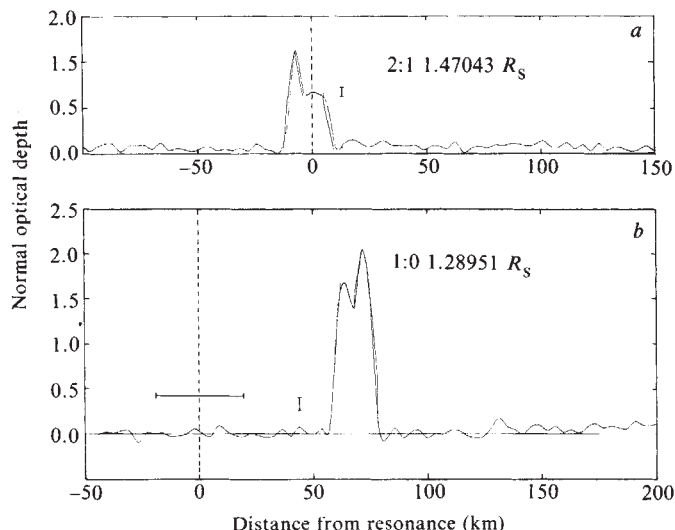


Fig. 4 The normal optical depths for regions located near the 1980S27 2:1 (a) and Titan apsidal resonances (b). The origins of the horizontal scales correspond to the expected location of each resonance. The horizontal error bars corresponding to the Titan apsidal resonance location are due to the uncertainty in Saturn's zonal harmonics¹⁵. The vertical error bars in each panel show the single sample noise level at the indicated τ .

is the distance from the resonance¹⁴. A more severe problem would develop from changing σ if we were to try to determine viscosity, because the wave amplitude can change significantly with changes in σ .

With these cautions in mind we proceed to determine σ and the opacity K via the following relations:

$$\sigma (\text{g cm}^{-2}) = \frac{0.325(m-1)\lambda d}{R^4} \quad (1)$$

$$K (\text{cm}^2 \text{g}^{-1}) = \frac{\tau}{\sigma} = \frac{R^4}{0.325(m-1)d\lambda} \frac{\tau}{\lambda} \quad (2)$$

where R is the resonance location in Saturn radii, and the wavelength λ and the distance from the wave source d are measured in kilometres. We find that all three of the A ring density waves exhibit strong enhancements of local (wavelength-averaged) τ over the initial 100 km of each wave train, indicating that density waves cause changes in local σ or K . We therefore analysed our data separately assuming constant α and constant K . The assumption of constant σ (equation (1)) does not provide an adequate fit to the observed wavelengths as a function of d . The assumption of constant K does, however, give a satisfactory fit to the data. On this assumption, the quantity τ/λ in equation (2) is determined from the observed wavelength-averaged τ for each wave. The A ring results for K and σ given in Table 1 and the wave source locations shown in Figs 2 and 3 are based on the constant opacity assumption.

Ring boundaries and ring mass

As we have noted previously, the 7:6 resonance is in extreme proximity to the outer edge of the A ring (Fig. 2). We feel it is not chance that the second strongest resonance in the rings should coincide with the outer edge of the ring system while the strongest resonance, the Mimas 2:1, lies at the outer boundary of the B ring. Satellites exert a torque on particles at resonances, removing angular momentum from them and thus pushing the particles closer to Saturn. A similar satisfactory explanation of the inner boundaries of the major rings is not yet available.

A detailed study of the outer edge of the A ring requires azimuthal coverage available only from imaging data. We would expect a seven-lobed pattern in analogy with the two-lobed

Table 2 Ring mass

Ring	Boundaries (R_s)	Mass ($10^{-8} M_s$)
C	1.2342–1.5259	0.2
B	1.5259–1.9477	5.0
Cassini division	1.9477–2.0232	0.1
A	2.0232–2.2676	1.1
Total		6.4

oval pattern seen at the Mimas 2:1 resonance. Observation of more frequently spaced undulations in the outer edge of the A ring boundary would be indicative of the effects of the higher m -order resonances present in the region (Fig. 2).

Our determinations of surface mass density at three points in the A ring constitute the first such measurements for this ring. Our single mass density measurement in the B ring at the 1980S1 2:1 resonance is in good agreement with the 60 g cm^{-2} at this resonance obtained earlier by Lane *et al.*⁶. All of our surface densities are consistent with a constant ring opacity of $0.01 \text{ cm}^2 \text{ g}^{-1}$, as are previous measurements^{5,6}. Using this result, together with our optical depth atlas, we can estimate ring masses as follows:

$$M_{\text{ring}} = 2\pi \int_{R_i}^{R_o} \sigma r dr = 2\pi \int_{R_i}^{R_o} \frac{\tau r dr}{K} = \frac{2\pi}{K} \int_{R_i}^{R_o} \tau r dr \quad (3)$$

where M_{ring} is the total mass of the ring, r is a radius in the ring and R_i and R_o are the inner and outer ring boundaries, respectively. The results of these integrations for the major rings are given in Table 2, together with UV spectrometer measurements of the ring boundaries. Our determination of the B ring mass may be underestimated due to the fact that we have assumed a maximum optical depth of $\tau = 2.85$ over the 15% of the B ring where we see no stellar signal. Measurements from other Voyager instruments may help to establish accurate optical depths in these regions. The A ring may also be more massive if the opacities we determine at the resonances are higher than the rest of the A ring due to an enhancement of small particles produced by collisions in the shocked region of the density waves.

Received 5 March; accepted 29 March 1982.

We find the total mass of the ring system to be $6.4 \times 10^{-8} M_s$, approximately the same mass as Mimas. Assuming uniform ring opacity, this total mass could be in error by perhaps 30%. The best previous estimate of the mass of the rings was the upper limit of $170 \times 10^{-8} M_s$ from the Pioneer 11 gravity results¹⁵. Our much smaller measurement indicates that an actual measurement of ring mass is probably unobtainable from existing spacecraft gravity measurements.

Conclusions

We have shown that resonances are associated with a wide range of morphological features in the rings. In low optical depth regions such as the C ring and inner Cassini division they are linked with gaps containing narrow double-peaked ringlets, while in higher τ regions they appear as density waves and wave-like disturbances. The two strongest resonances are coincident with the outer edge of the B ring and the outer edge of the A ring. Known resonances have a major role in determining the structure of the A ring but can explain only a small fraction of the observed features in the B and C rings, where resonances are less frequent. We have analysed the four density wave trains to determine their point of origin and the local surface mass density. These estimates of surface mass density in combination with our optical depth atlas make possible an estimate of the total mass and mass budget of the ring system.

We thank J. N. Cuzzi and F. H. Shu for helpful discussions. This work was supported by the JPL, California Institute of Technology under NASA contract NAS 7-100. Additional support was provided by the Planetary Sciences Discipline of NASA's Office of Space Sciences grant NAGW-62. J.J.L. was supported by NASA training grant NGT-05-003-800.

7. Goldreich, P. & Tremaine, S. *Icarus* **34**, 240 (1978).
8. Lissauer, J. J. & Cuzzi, J. N. *Astr. J.* (in the press).
9. Broadfoot, A. L. *et al. Space Sci. Rev.* **21**, 183 (1977).
10. Sandel, B. R. *et al. Science* **215**, 548 (1982).
11. Dermott, S. F. & Murray, C. D. *Icarus* **48**, 12 (1981).
12. Elliot, J. L. *et al. Astr. J.* **86**, 444 (1981).
13. Shu, F. H., Milione, V. & Roberts, W. W. *Astrophys. J.* **183**, 819 (1973).
14. Shu, F. H. *Astrophys. J.* **160**, 99 (1970).
15. Null, G. W. *et al. Astr. J.* **86**, 456 (1981).

1. Collins, S. A. *et al. Nature* **288**, 439 (1980).
2. Smith, B. A. *et al. Science* **212**, 163 (1981).
3. Smith, B. A. *et al. Science* **215**, 504 (1982).
4. Franklin, F. A., Colombo, G. & Cook, A. F. *Nature* **295**, 128 (1982).
5. Cuzzi, J. N., Lissauer, J. J. & Shu, F. H. *Nature* **292**, 703 (1981).
6. Lane, A. L. *et al. Science* **215**, 537 (1982).

Stratospheric condensation nuclei variations may relate to solar activity

D. J. Hofmann & J. M. Rosen

Department of Physics and Astronomy, University of Wyoming, Laramie, Wyoming 82071, USA

Observations of increases of stratospheric condensation nuclei suggest a photo-initiated sulphuric acid vapour formation process in spring in polar regions. It is proposed that the sulphuric acid rapidly forms condensation nuclei through attachment to negatively charged multi-ion complexes and that the process may be modulated through variations in solar activity.

THE possible connections between solar variability and the weather on the one hand and climate on the other have been widely debated^{1,2}. In studies of the correlation of solar activity with both weather and climate, there has been little or no experimental evidence to show by which mechanism minute changes in solar energy incident at the Earth's orbit may be transformed to observed weather or climate variations. A direct coupling through variations in the solar constant has generally been discarded in favour of a triggering mechanism which utilizes charged solar corpuscular radiation. The latter may take the form of solar control of galactic cosmic radiation through the waxing and waning of solar magnetic irregularities in the interplanetary medium. The triggering mechanism may also appear as direct magnetospheric intrusion of solar flare accelerated charged particles and/or energetic electrons released from

the magnetosphere by enhanced geomagnetic activity due to active Sun processes.

The intermediary in the solar variation-weather connection may be the stratospheric ozone layer which can be depleted by the production of nitric oxide during charged particle bombardment (solar or galactic) of the upper atmosphere and thereby may affect the intensity of solar UV radiation in the Earth-atmosphere system. It may also take the form of the creation of atmospheric ionized species through charged particle bombardment which alter the electrical characteristics of the atmosphere and thereby modulate the occurrence of thunderstorms.

One of the more effective ways of altering climate, variations in cloudiness or the particulate content of the atmosphere, has perhaps the least amount of observational support as far as solar activity correlations are concerned. Increased strato-

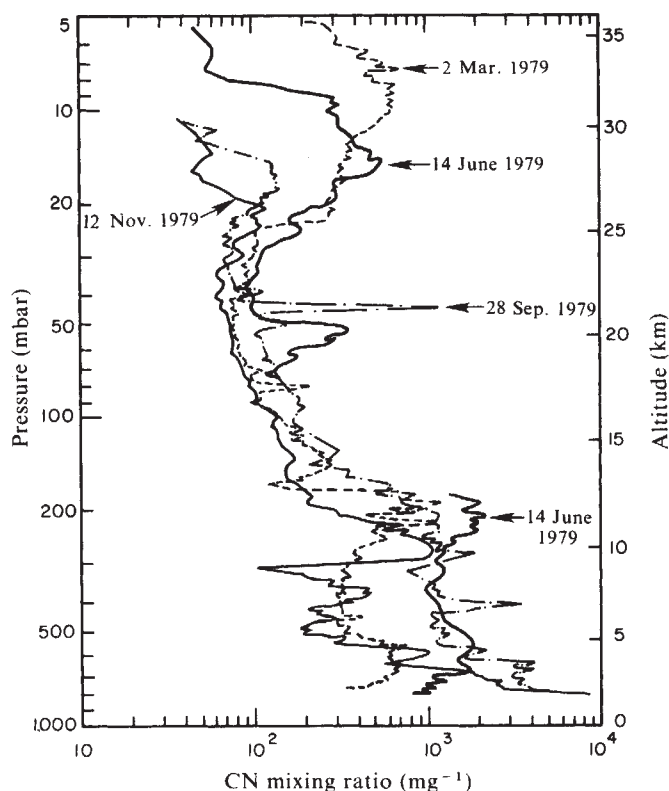


Fig. 1 Condensation nuclei ($r \geq 0.01 \mu\text{m}$) mixing ratio profiles during 1979 showing the time variation of the high altitude event.

spheric particulates, due to volcanic eruptions, are capable of heating the stratosphere and cooling the surface. Thus, any mechanism which can create material particles in the atmosphere is a likely candidate for climatic studies.

Ney³ suggested that the logical place to look for solar-weather effects was the solar modulation of galactic cosmic radiation. Later Roberts and Olson⁴ proposed that atmospheric ions may serve as condensation nuclei (CN) for the formation of cirrus clouds in the 10–12 km region, thus forming a solar activity connection with climate. Dickinson⁵ proposed that ions formed by the galactic cosmic radiation may be important in the formation of stratospheric sulphuric acid aerosol which, in turn, could nucleate cloud droplets in the upper tropospheric environment. With primary ionization control ascribed to the galactic cosmic radiation, minimal ion density at solar maximum would suggest fewer growth sites, less competition for the available sulphuric acid vapour and thus larger though fewer particles at this time.

The formation of stratospheric aerosol through ions has largely been neglected. Thus, the stratospheric aerosol formation models of Rosen *et al.*⁶ and Turco *et al.*⁷ included only a tropospheric source of pre-existing CN which diffused into the stratosphere and served as sites for larger aerosol particle growth by accretion of sulphuric acid vapour. This assumption of a tropospheric source of CN was justified by the measurements^{8,9} wherein the CN mixing ratio peaked just below the tropopause and fell off sharply in the stratosphere. However, the observation of an essentially constant, albeit low, mixing ratio from 20 km to the highest altitude of observation (~32 km) left some room for speculation that a high altitude source (for example, meteoritic) may be operative.

The possibility that this source could be ionic in nature was not considered because even at typical stratospheric ion concentrations of $5,000 \text{ cm}^{-3}$, ion-induced nucleation cannot compete with heterogeneous nucleation on pre-existing particles for typical stratospheric particle concentrations¹⁰ and only following major solar flare particle events (which are very rare), can ion induced nucleation be important.

Recent observations have cast new light on this problem. Following the initial stratospheric ion-mass spectrometer measurements¹¹, Ferguson¹² interpreted the results as indicating the clustering of ions, forming stable ion pairs, and suggested their possible importance in nucleation processes in the stratospheric sulphate layer.

Arnold and Hensen¹³ followed the initial positive ion measurements with measurements of negative ion clusters which suggested the presence of sulphuric acid in these clusters. Arnold¹⁴ suggested that attachment of free ions to these clusters could lead to larger multi-ion complexes (MIC), or that they may even mutually coalesce. He suggested that formation of MIC and subsequent growth could result in a gas-to-particle conversion mechanism which does not require high gas supersaturations. In the presence of a supersaturated gas, such as sulphuric acid vapour, MIC which have grown to embryonic size could act as condensation nuclei. Arnold and Hensen¹⁵ found a pronounced layer of MIC at 30 km having masses in excess of 327 AMU and labelled them 'pre-condensation nuclei'. These new observations have prompted us to examine some of our recent stratospheric particle observations, the results indicating that a high altitude ion-CN relation may exist.

Observations

We have recently reported¹⁶ the observation of an unusual increase in the CN ($r \geq 0.01 \mu\text{m}$) concentration above 25 km which lasted about 6 months in 1979 (March through September) and some evidence for its repetition in March, 1980 at Laramie (41°N). Measurements were made using a balloon-borne thermal growth chamber, attached to a light scattering particle counter. In 1979 the increase was peaked at ~32 km with a mixing ratio of $\sim 650 \text{ mg}^{-1}$ above a normal background of $\sim 30 \text{ mg}^{-1}$ at this altitude (a concentration of $\sim 9 \text{ cm}^{-3}$ above a background of $\sim 0.4 \text{ cm}^{-3}$).

No attempt was made to explain the nature of this event as it was not conclusive if the feature was really annual on the basis of the data at hand. Earlier measurements probably would not have detected such an annual event, if it existed, because the CN growth chamber only functioned satisfactorily to an

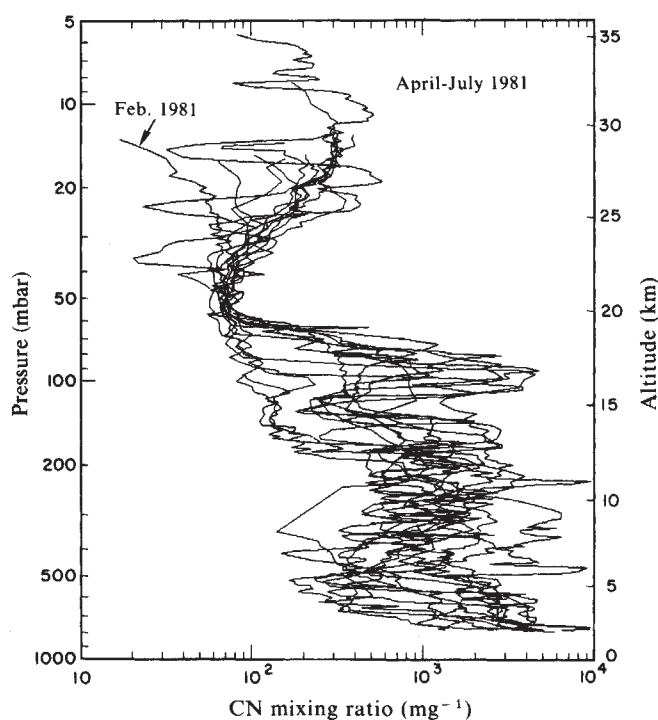


Fig. 2 Condensation nuclei ($r \geq 0.01 \mu\text{m}$) mixing ratio profiles during 1981 showing the high altitude event. Enhanced CN layers in the 13–18 km region were due to the volcanic eruption of Alaid on 28 April 1981.

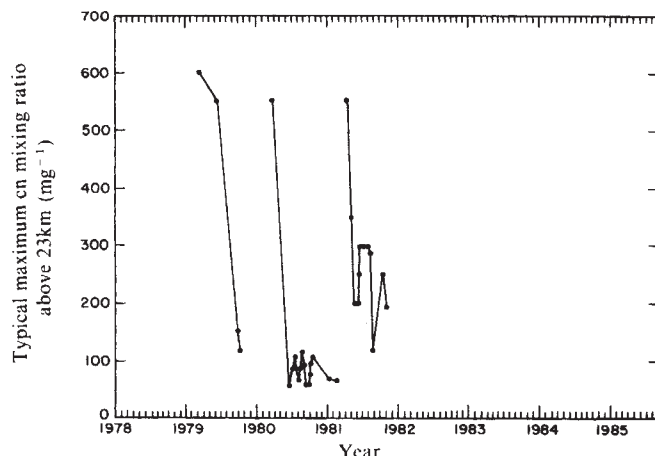


Fig. 3 Condensation nuclei ($r \geq 0.01 \mu\text{m}$) time variation showing the occurrence of the springtime events at Laramie.

altitude of 25 km (refs 8, 9). The 1980 investigations were complicated by the eruption of Mt St Helens in May 1980 and while numerous measurements were made, smaller balloons were used to study the 20–25 km effects of Mt St Helens and the 30–35 km region was not effectively studied until September when CN levels appeared normal¹⁷.

Five soundings during 1979, shown in Fig. 1, indicated that the excess CN decayed exponentially with a e^{-1} time of 130 days. As the layer decayed, the mixing ratio peak appeared at successively lower altitudes, finally blending into the background at about 25 km. Initial attempts to explain the event as being due to the burnup of a large piece of space junk in the atmosphere were rapidly discarded. One tonne of such material, distributed uniformly in a 10-km thick cloud with a particle concentration of 10 cm^{-3} (mixing ratio of 600 mg^{-1} at 30 km) would have a horizontal extent of only 38 km for particle sizes of $0.07 \mu\text{m}$ (typical CN sizes in the unperturbed stratosphere) and 700 km for sizes of $0.01 \mu\text{m}$, a lower limit of detection for the instrument. Such a small cloud would not have been detected on all five CN flights conducted in 1979. Considerations of a large meteor would result in the same conclusions. The observation that the CN mixing ratio displays a well defined maximum between 25 and 35 km further rules out objects entering the atmosphere from space. The latter, if observable, would display a broad mixing ratio maximum extending to great altitudes. It was considered more likely that the layers observed in 1979 were all derived through *in situ* production at their origin but that variable transport from a source region resulted in the observed characteristics.

In 1981 the high altitude enhancement did not seem to be present in February, at least in the 25 km region, but it was very apparent in April and persisted through 12 subsequent soundings. The data are shown in Fig. 2. Some complications resulted from the volcanic eruption of Alaid in the Kurile Islands in April 1981 but these disturbances were restricted to the 13–18 km region¹⁸. Figure 3 shows the high altitude CN mixing ratio versus time and clearly demonstrates the periodic nature of the events.

Discussion

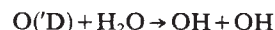
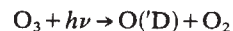
The observation that the CN events are seasonal, favours variable transport to midlatitudes from a remote source region as there is no known seasonally varying production source at midlatitude. Little is known about the nature of meridional transport or its seasonal variations at altitudes in the 30 km region. If one associated possible meridional transport with regions of high wind speed, then likely candidates at 30 km in the Northern Hemisphere are the polar vortex region at 50–80° N during December–February ($v = 20\text{--}40 \text{ m s}^{-1}$) and the region 0–30° N during June–August ($v = 20 \text{ m s}^{-1}$) (ref. 19). In view of our observation that the enhanced CN layers appear

in March, the polar regions appear to be the most promising. In addition, soundings in Panama (9° N) in 1976 and 1977 and in Brazil (5° S) in 1978 did not reveal excess CN mixing ratios in the upper stratosphere although the CN detector used then was suspect above 25 km (ref. 9).

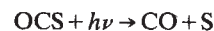
In view of the more probable source region for the CN enhancements being at high latitude, and the fact that multi-ion complexes have been observed in a layer at the same altitude as the CN enhancements appear, we feel that there may be a logical connection between the two and thus possibly a connection between solar activity and stratospheric aerosol formation.

The enhancements are not observed until March at the earliest while the polar vortex 30 km wind system is operative throughout winter. This suggests that excess CN may not be formed at 30 km during the polar winter night but begin to form in spring. Thus a sunlight effect in the formation of CN appears to be a possible mechanism.

As large aggregate negative ion clusters contain sulphuric acid, and since stratospheric CN is largely volatile at 150 °C, suggesting a sulphuric acid composition¹⁶, it seems possible that a sunlight effect may be manifested through the photochemical production of sulphuric acid. Thus we propose that production of the acid is limited during the winter night while sulphur-bearing precursor gases such as SO_2 , following volcanically active periods, or OCS, during background conditions²⁰, are transported poleward and build up along with O_3 . Sulphuric acid may be formed following the oxidation of SO_2 either by OH or O. For relative humidities $>1\%$, OH formed through the photodissociation of O_3



is the dominant oxidizing agent while at lower relative humidities, ground state atomic oxygen, formed through the photodissociation of O_2 and O_3 , is dominant²¹. In addition, because at least one of the initial steps in the conversion of OCS to sulphuric acid in the stratosphere involves photolysis



it appears likely that sunlight reaching the polar stratosphere

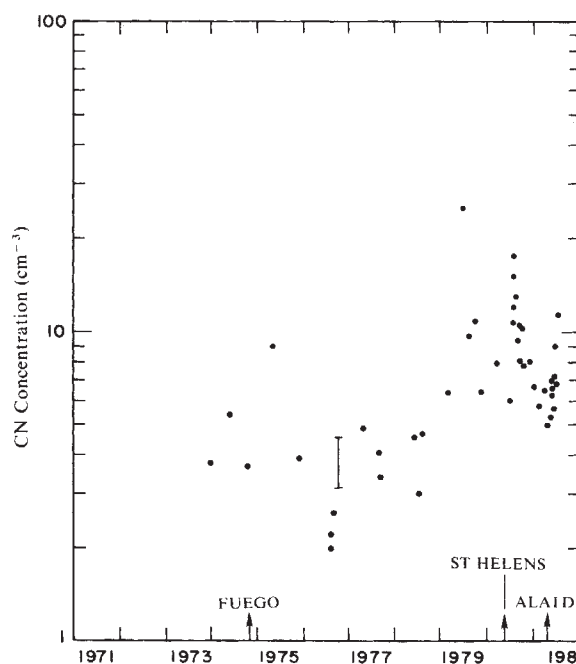


Fig. 4 The condensation nuclei concentration at 20 km as a function of time since measurements began at Laramie in 1973. The vertical bar is the range of concentration observed in four soundings over a 12-day period. The arrows indicate major volcanic eruptions which perturbed the stratospheric $r \geq 0.15 \mu\text{m}$ particle concentration.

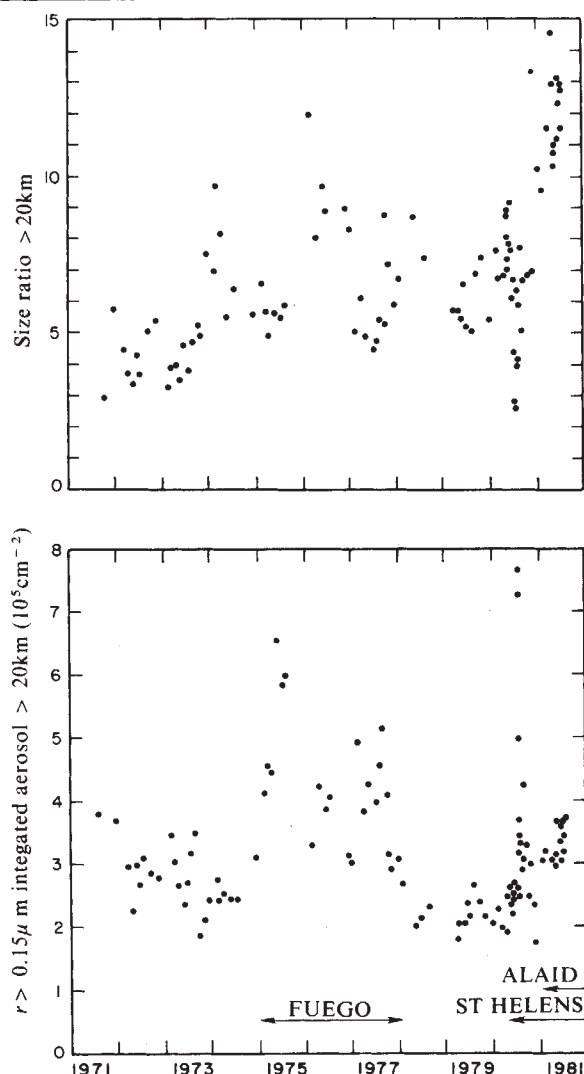


Fig. 5 $r \geq 0.15 \mu\text{m}$ particle loading and the ratio of $r \geq 0.15 \mu\text{m}$ to $0.25 \mu\text{m}$ particle loading above 20 km as a function of time since measurements began at Laramie in 1971. Periods when the stratospheric aerosol was perturbed by volcanic eruptions are indicated.

in spring would enhance the production of H_2SO_4 and allow ion complexes to grow rapidly into CN and subsequently be transported to middle latitudes beginning in March.

As the formation of CN in this model is related to multi-ion complexes, it may be related to solar activity which modulates the atmospheric ion density through modulation of the galactic cosmic radiation intensity, large solar flare accelerated particle events, and enhanced auroral activity or precipitation of energetic electrons during geomagnetic activity.

To investigate these speculations, measurements of MIC, CN and larger aerosol at 30 km are required over a solar cycle. Our measurements of CN date back to 1973 and of larger aerosol ($r \geq 0.15 \mu\text{m}$) to 1971. Balloon altitudes in the 27–29 km range were generally attained. Volcanic effects can be large in the 20-km region especially for the larger aerosol. CN are also affected by volcanic eruptions, but these effects are comparatively shortlived due to rather rapid coagulation of these small particles with themselves and with the larger aerosol. While the CN detector functioned satisfactorily only to ~25 km before 1979, CN data at 20 km are reliable and relatively free of volcanic perturbations.

Thus, of the data at hand, it would appear most fruitful to examine the time variation of the CN concentration at 20 km

for solar cycle effects. These data are shown in Fig. 4. There seems to be an increasing trend which was evident even before the eruption of Mt St Helens which severely disturbed stratospheric CN concentrations from June to August 1980 at 20 km. The higher concentrations seem to occur at the time of maximum solar activity. In fact, the increasing trend may simply be the result of an accumulation of high altitude events over the past several years. Most of the increase has taken place since early 1979; sunspot maximum occurred in December 1979.

Thus there seem to be more particles at solar maximum, at least for the CN size range. We can investigate the size aspect for the larger aerosol from $r \geq 0.15 \mu\text{m}$ measurements. By measuring two sizes ($r \geq 0.15$ and $0.25 \mu\text{m}$), we can calculate a size ratio (ratio of the concentrations of the two sizes) which will vary as the size distribution varies. Statistics are poor above 25 km, but if one integrates the concentration above 20 km, a reasonable set of data can be obtained. Figure 5 shows integrated concentration or particle loading (particles cm^{-2}) above 20 km for the $r \geq 0.15 \mu\text{m}$ particles and the ratio of loading for $r \geq 0.15 \mu\text{m}$ to $r \geq 0.25 \mu\text{m}$ for all soundings which reached altitudes in excess of 27 km.

The period 1975–77 was severely disturbed at 20 km by the volcanic eruption of Fuego²² and the period after May 1980 by the Mt St Helens eruption¹⁷ and the Alaid eruption in April 1981¹⁸. The period 1978–early 1980 appears to have been free from volcanic eruption.

Sunspot numbers were about 40 in 1973 compared with 140 in 1979 so that the 1972–74 data would represent solar inactive conditions and the 1978–80 data an active solar period. The data in Fig. 5 do not indicate sizable trends between these periods. The size ratio suggests a general trend towards smaller particles (a higher ratio) during the active period. This would accord with the observed increase in condensation nuclei over the period and thus refute Dickinson's hypothesis that ionization control of CN formation could be through the galactic cosmic radiation.

While volcanic effects seem to muddle the interpretation, they may also be used to clarify stratospheric conditions for particle growth between periods. A noticeable effect has been that particles which formed in the stratosphere following the eruption of Fuego in 1974 were substantially larger than those which formed following the eruption of La Soufriere and Sierra Negra in 1979, St Helens in 1980 and Alaid in 1981. The ratio of $r \geq 0.15 \mu\text{m}$ to $r \geq 0.25 \mu\text{m}$ particle concentrations at the stratospheric maximum was typically 3–4 following the Fuego eruption, 5–6 at background in 1978–79 and 6.5–7.5 following the La Soufriere–Sierra Negra, Mt St Helens and Alaid eruptions. Thus the data are consistent with smaller particles forming during solar maximum periods.

The parameters which control the size to which particles grow in the stratosphere are the amounts of H_2SO_4 and H_2O vapours available and the number of condensation sites present. While the Fuego eruption may have injected relatively more vapours than the other eruptions, it is of interest that the background CN levels were considerably higher at the time of the recent eruptions compared with the time of the Fuego eruption. Higher levels of condensation sites would result in generally smaller particles due to increased competition for the available sulphuric acid and water vapour. Thus, if increased ionization activity during solar maximum results in enhanced levels of ion complexes which form CN, then one would expect volcanically created particles to be smaller during solar maximum, as has been observed.

Conclusions

If solar activity directly affects the stratospheric aerosol, then this influence is probably most readily observable in the number of CN present in the upper stratosphere so that continued monitoring of this important constituent at the highest altitudes is imperative. It appears that a possible mechanism, which

relates ionization produced during solar active periods to the production of CN, now exists and can be used as a basis for further study of this possibly climatic related phenomenon.

We thank the Atmospheric Physics staff and students including J. Anderson, J. Harder, N. Kjöme, D. Martell, G. Olson

Received 29 December 1981; accepted 16 March 1982.

1. Herman, J. R. & Goldberg, R. A. *Sun, Weather and Climate*, NASA SP-426 (NASA, Washington DC, 1978).
2. McCormac, B. M. & Seliga, T. A. *Solar-Terrestrial Influences on Weather and Climate* (Reidel, Boston, 1979).
3. Ney, E. P. *Nature* **183**, 451 (1959).
4. Roberts, W. O. & Olson, R. H. *Rev. Geophys. Space Phys.* **11**, 731-740 (1973).
5. Dickinson, R. E. *Bull. Am. Met. Soc.* **56**, 1240-1248 (1975).
6. Rosen, J. M., Hofmann, D. J. & Singh, S. P. *J. Atmos. Sci.* **35**, 1304-1313 (1978).
7. Turco, R. P., Hamill, P., Toon, O. B., Whitten, R. C. & Kiang, C. S. *J. Atmos. Sci.* **36**, 699-717 (1979).
8. Rosen, J. M. & Hofmann, D. J. *J. appl. Met.* **16**, 56-62 (1977).
9. Rosen, J. M., Hofmann, D. J. & Kaselau, K. H. *J. appl. Met.* **17**, 1737-1740 (1978).

and J. Vali for assistance with the numerous balloon soundings. This work was supported jointly by the NSF (Atmospheric Sciences Section), NASA (Ames Research Center), the NOAA (Aeronomy Laboratory) and the Office of Naval Research (Atmospheric Science Program).

10. Mohnen, V. A. *Proc. 4th CIAP Conf.*, 478-491 (US Department of Transportation, DOT-TSC-OST-75-38, 1975).
11. Arnold, F., Krankowsky, D. & Marien, K. H. *Nature* **267**, 30 (1977).
12. Ferguson, E. E. *Geophys. Res. Lett.* **5**, 1035-1038 (1978).
13. Arnold, F. & Henschen, G. *Nature* **257**, 521 (1978).
14. Arnold, F. *Nature* **284**, 610-611 (1980).
15. Arnold, F. & Henschen, G. *Geophys. Res. Lett.* **8**, 83-86 (1981).
16. Rosen, J. M. & Hofmann, D. J. *Rep. No. AP-61* (Department of Physics and Astronomy, University of Wyoming, 1981).
17. Hofmann, D. J. & Rosen, J. M. *Rep. No. AP-63* (Department of Physics and Astronomy, University of Wyoming, 1981).
18. Hofmann, D. J. & Rosen, J. M. *Geophys. Res. Lett.* **8**, 1231-1234 (1981).
19. Newell, R. E. *Scient. Am.* **224**, 32-42 (1971).
20. Hofmann, D. J. & Rosen, J. M. *J. Atmos. Sci.* **38**, 168-181 (1981).
21. Friend, J. P., Barnes, R. A. & Vasta, R. M. *J. phys. Chem.* **84**, 2423-2436 (1980).
22. Hofmann, D. J. & Rosen, J. M. *J. geophys. Res.* **82**, 1435-1440 (1977).

Expression of two proteins from overlapping and oppositely oriented genes on transposable DNA insertion element IS5

Bodo Rak, Monika Lusky* & Majka Hable

Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg, FRG

IS5, the 1,195 base pair transposable DNA element of Escherichia coli, contains two genes. One gene, which encodes a protein with an apparent molecular weight of 12,300, is, together with its promoter, completely contained within a larger gene which encodes a protein of molecular weight 37,000. The two genes are transcribed in opposite directions.

TRANSPOSABLE DNA elements have been found in prokaryotic and higher cells (for recent reviews, see refs 1-3). Although there is a large body of information on structural and genetic properties of these elements, the process of transposition itself remains to be elucidated. The differences in the structural organization of various transposable elements may be an indication of differences in at least some of the reaction steps involved in transposition. Several models of transposition propose a two-step reaction^{4,5}. In the first step, a transposition intermediate with co-integrated donor and recipient DNAs is formed (provided both molecules involved are circular DNA molecules). Co-integrated molecules have been detected in the cases of Tn3⁶, IS1⁷, IS903⁸ and several others. In the second step, the co-integrated intermediate is resolved. General recombination or, as has been shown for Tn3⁹, site-specific recombination in recombination-deficient (*recA*) cells, can perform this step. This pathway of transposition is now widely accepted for Tn3, although it is still unclear for other elements. In the 5,000 base pair (bp) element Tn3, the region carrying the functions required for transposition is at least 3,800 bp long and contains two genes plus an internal resolution site (IRS) for site-specific recombination^{6,9-11}. Tn501 and Tn1721^{12,13} appear to be similarly organized and, together with Tn3 and several others, constitute one class of large transposable elements. A different model for transposition, not requiring a co-integrated structure as an intermediate, has recently been proposed for the transposable element *mu* (the 38,000 bp genome of bacteriophage Mu) on the basis of electron microscopic observations¹⁴.

An additional class of transposable elements is represented by the insertion (IS) elements, the first transposable elements to be discovered in *Escherichia coli*^{15,16}. They are considerably

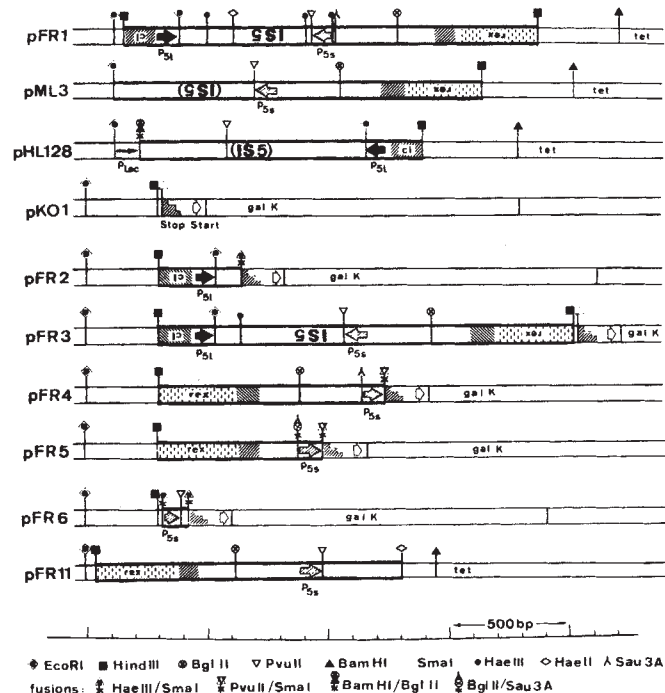
smaller (800-1,500 bp) and do not carry any detectable phenotype although they are autonomously transposable. The mechanism of transposition of IS elements remains to be elucidated, but their genetic organization is expected to differ from that of the large transposable elements of the Tn3 type mainly because their coding capacity is restricted. Recent determinations of the DNA sequences of several insertion elements (IS1, IS2, IS4, IS5, IS903 and IS102)^{8,17-23} have revealed an open reading frame extending over most of the elements' lengths. A corresponding protein has only been detected for IS4²⁴. Interestingly, all these elements also possess a second, shorter reading frame of significant length and totally contained within the first one. This reading frame is in opposite orientation to the longer one and has the same codon-codon register. Here we present evidence for the simultaneous expression of two proteins from both opposing reading frames on IS5 under the control of element-coded transcription and translation start signals.

Two promoters on IS5 precede the large and small reading frames

Analysis of the DNA sequence of insertion element IS5²⁰⁻²² revealed two overlapping reading frames in opposite orientations. If both reading frames represent functional genes they must be preceded by start signals for transcription and translation. To test IS5 for potential promoters we cloned different IS5 DNA fragments in front of a structural gene whose gene product could easily be measured and which did not possess a promoter. Plasmid vector pKO1 (ref. 25), which carries the galactokinase structural gene without a preceding promoter sequence, was used. Readthrough of ribosomes from any gene that may be inserted proximal to the galactokinase gene is prevented in this vector by nonsense codons in all three reading frames. DNA sequences inserted into restriction sites in front of the galactokinase gene could thus be tested for promoter

* Present address: Department of Molecular Biology, University of California, Berkeley, California 94720, USA.

Fig. 1 Relevant structures and restriction sites of plasmids used. Note that the right arm of IS5 is defined as the one carrying the *EcoRI* recognition site. A 1,720 bp *HindIII* fragment including the entire IS5 element²² was prepared from λ KH100:IS5 DNA²⁶ and ligated into the *HindIII* site of the vector pML2 to yield pFR1. Plasmid vector pML2 is a derivative of pBR322 in which a DNA fragment from position 1,095 to 2,485 (ref. 46) is deleted (ref. 47 and M.L., unpublished). Plasmid pML3 was obtained by *in vitro* deletion of the small *EcoRI* fragment of pFR1. The construction of pHL128 has been described earlier³². Plasmid pKO1 (ref. 25) consists of a segment of pBR322 DNA from the *PvuII* restriction site at position 2,067 (which is destroyed) to the *EcoRI* site at position 4,361 plus the galactokinase gene. Single restriction sites for *EcoRI*, *HindIII* and *SmaI* are located proximal to the galactokinase gene, followed by translation stop codons in all three reading frames (■). A translation start sequence precedes the galactokinase gene (⇒). There is no proximal promoter sequence that could give significant transcription of the galactokinase gene. Plasmid pFR2 was constructed by substituting the 17 bp *HindIII*-*SmaI* fragment of pKO1 with the 357 bp *HindIII*-*HaeIII* fragment of pFR1 containing the right end of IS5 up to the first *HaeIII* site (IS5: 1,195–980, IS5 sequence coordinates as in refs 20, 21). Plasmid pFR4 was similarly constructed, but the *HindIII*-*PvuII* fragment of pFR1 containing the left end of IS5 (IS5: 1–517) was used as the substituting fragment. For the construction of pFR3, pKO1 was cut with *HindIII*, treated with calf intestinal alkaline phosphatase and ligated with the *HindIII* fragment of pFR1, which contained the entire IS5 element. To construct pFR5, plasmid pFR1 DNA was digested with *HindIII*, *BglII* and *PvuII* and the resulting fragments were separated on a 1.2% agarose gel. The 585 bp *HindIII*-*BglII* fragment and the 359 bp *BglII*-*PvuII* fragment were eluted⁴⁸, and the 359 bp fragment was digested with *Sau3A*, resulting in two fragments of 251 and 108 bp respectively. The fragments were separated on a 7.5% polyacrylamide gel and the 108 bp *Sau3A*-*PvuII* fragment was eluted⁴⁹. It was then mixed with the 585 bp *HindIII*-*BglII* fragment and pKO1 DNA deleted for the 17 bp *HindIII*-*SmaI* fragment and the mixture was ligated. Lysates of independent transformants⁵⁰ were analysed for the structures of the plasmids present using *Sau3A* digests as well as *HindIII*-*TaqI* and *HindIII*-*HaeII* double digests. To construct pFR6, a 975 bp *HindIII*-*BstEII* fragment (which contains the left end of IS5 up to IS5:548; the *BstEII* site lies 8 bp to the right of the *HaeIII* site at IS5:540, not marked in the figure) was isolated from pFR1 DNA. This DNA fragment was digested with *HaeIII* and the resulting fragments separated on a 7.5% acrylamide gel and eluted. The 105 bp fragment bordered by the *HaeIII* sites at IS5:435 and IS5:540 was ligated into pKO1, cut with *SmaI* and treated with calf intestinal alkaline phosphatase. Small lysates from individual transformants were analysed by double digestion with *EcoRI* and *PvuII* and with *TaqI* and *PvuII*. Plasmid pFR11 was constructed as follows. (1) The IS5-containing *HindIII* fragment of pFR1 was inverted by digestion of pFR1 DNA with *HindIII* and re-ligation. (2) The plasmid containing the inversion (pFR8) was digested with *BstEII* and *BamHI*, and both resulting fragments separated on a 0.8% agarose gel and eluted. (3) The smaller of the two fragments, containing IS5:543–1,195, the *AcI* part plus pBR322 sequences of pML2 from position pBR322:29 to position 375 (ref. 46), was restricted with *HaeII* and fractionated on a 1.8% agarose gel. (4) The *BstEII*-*HaeII* fragment (IS5:543–838) and the *HaeII*-*BamHI* fragment (pBR322:236–375) were eluted, ligated and then digested with *BstEII* and *BamHI* to eliminate head-to-head ligation products. (5) The large *BstEII*-*BamHI* fragment containing the pML2 moiety (see step 2) was added together with additional ligase. Small lysates were made from individual transformants and identified by restriction with endonucleases *BamHI*, *BstEII* and *EcoRI*. (6) One plasmid (pFR10) carried a deletion of pFR8 from the *HaeII* site at IS5:838 to the *HaeII* site at pBR322 coordinate 236 of vector pML2. (7) Since pFR10 carries a deletion of the C-terminal part of the tetracycline gene, the small *EcoRI*-*BamHI* fragment of pBR322 was substituted for the small *EcoRI*-*BamHI* fragment of pFR10 giving rise to pFR11. The structure of pFR11 was verified by analysis with restriction endonucleases *EcoRI*, *BstEII*, *PvuII*, *HindIII*, *BamHI*, *PstI* and *HaeII*. Plasmid pFR11 contains the promoter, translation start and 104 amino acids of the IS5 small reading frame plus leader and coding sequence of the larger tetracycline gene which, when transcribed and translated, would give a protein of 450 amino acids.



activity. Figure 1 shows the pertinent structures of vector pKO1 and its derivatives harbouring different IS5 DNA fragments. Table 1 shows the Gal phenotypes on indicator plates of *galK*⁻ strains harbouring these plasmids, as well as the amounts of galactokinase expressed by the various plasmids *in vivo* and in a cell-free system. Nucleotide map positions are those given in ref. 20. Orientation of IS5 is defined with respect to its integration site in the *cI* gene of bacteriophage λ in mutant λ KH100 (ref. 26), and its short *EcoRI* arm (right arm) points towards the beginning of gene *cI*. The source of the different subfragments inserted into pKO1 was pFR1 (Fig. 1), which contains the 1,195 bp IS5 element on a 1,720 bp *HindIII* fragment derived from λ KH100 DNA.

Plasmid pFR2 contains the right arm of IS5 up to the *HaeIII* site at position IS5:980 inserted in front of the galactokinase gene as a *HindIII*-*HaeIII* fragment. The data presented in Table 1 show that a promoter of moderate strength active both *in vivo* and in the cell-free system, is located on this cloned fragment. In addition to IS5 sequences (IS5:980–1,195), the *HindIII*-*HaeIII* fragment contains an internal segment of the *AcI* structural gene, λ :38,150–38,291 (G. Hobom, personal communication), known not to contain any promoter signals^{27,28}. We therefore conclude that a promoter, *p*₅₁, which transcribes towards the centre of IS5, is located on the right arm of the element. Deletion of the hybrid vector insert *EcoRI* fragment from pFR2 results in a Gal⁻ phenotype on transformation of a *galK*⁻ strain (pFR7, data not shown). This defines the location of *p*₅₁ between the *EcoRI* site on IS5 (IS5:1,097) and the end of the element (IS5:1,195). The localization of a functional promoter in this region agrees with the prediction

of a promoter (*p*₅₁) at IS5:1,175–1,139 from the sequencing data²⁰. This promoter precedes the long open reading frame of IS5. Galactokinase activity was significantly reduced when all of IS5 was inserted proximal to the galactokinase gene in the same orientation as the IS5 fragment in pFR2 (pFR3). This reduction in expression of *galK* may be due to partial termination of transcription at a terminator sequence, *t*₅₁, located on the left arm of IS5²⁰.

Cloning of the left arm of IS5 (as *HindIII*-*PvuII* fragment) into pKO1 in front of *galK* (pFR4; Fig. 1) resulted in synthesis of galactokinase in a cell-free system but not *in vivo* (Table 1). This rather low but significant promoter activity should be located in the IS5 segment of this fragment, because the λ segment also present, which contains the end of the *AcI* gene and the beginning of the *lexA* gene (λ :37,727–38,153; G. Hobom, personal communication) should not code, in this orientation, for any promoter reading towards the galactokinase gene^{27,28}. Deletion of the DNA between the *BglII* site at IS5:158 and the *Sau3A* site at IS5:409 within pFR4, giving rise to pFR5 (Fig. 1), did not abolish the observed promoter activity (Table 1).

The position of this promoter was mapped more precisely by constructing hybrid plasmids each carrying one of the different fragments that can be generated by *HaeIII* digestion of the 975 bp DNA segment between the *HindIII* site at λ :37,727 and the *BstEII* site at IS5:548. The fragments were inserted in front of the galactokinase gene in both possible orientations. Only one of the *HaeIII* fragments (IS5:435–540) was capable of promoting galactokinase synthesis in the cell-free system when inserted into pKO1 (pFR6, Table 1). Its

Table 1 Activity of promoter signals on IS5

Plasmid	Gal phenotype	Galactokinase (units)	
		Cell free	<i>In vivo</i>
pKO1	—	13.0	5.4
pFR2	+	82.4	91.5
pFR3	(+)	29.5	11.8
pFR4	—	35.2	1.5
pFR5	—	38.5	2.3
pFR6	—	46.6	4.8

E. coli strain N100 (*galK*, *recA*) was transformed with the different plasmids indicated and Gal phenotypes were detected on McConkey galactose ampicillin agar plates. N100/pFR2 colonies were immediately red on these plates whereas N100/pFR3 developed red colonies only after 1 day incubation at 37 °C. All other plasmids gave white colonies in N100. Galactokinase synthesis directed by the various plasmids was also measured in strain N100. Bacteria were grown to $A_{546} = 1.0$ in minimal salt medium supplemented with casamino acids and glucose, collected by centrifugation and frozen. Cell-free synthesis of galactokinase directed by the different plasmid DNAs was carried out as described by Zubay *et al.*⁴² with modifications⁴³. Synthesis was for 45 min with 40 $\mu\text{g ml}^{-1}$ of plasmid DNA. Galactokinase was assayed from the frozen cells or from the cell-free reaction mixture as described elsewhere^{25,43,44}. Galactokinase activity is expressed in nmol of ^{14}C -galactose phosphorylated per minute and per 1.0 A_{546} bacteria (*in vivo*) or per ml reaction volume (in the cell-free system). Galactokinase synthesis directed by plasmid pKO1 alone may be explained by readthrough from promoter p_a of pBR322, which is the most efficient promoter present on pBR322, located around position 2,900 in clockwise orientation⁴⁵. Data given represent the average of three or more independent measurements with variations not exceeding 3 units for the *in vivo* assays or 5 units for the cell-free assays.

orientation in this clone is the same relative to *galK* as to the small reading frame in IS5 (Fig. 1). This DNA fragment overlaps with the beginning of the IS5 small reading frame. Its promoter activity can, therefore, be correlated with a DNA fragment immediately preceding the IS5 small reading frame. Other hybrid plasmids directed galactokinase synthesis in this system to the level of pKO1 or below (data not shown).

The 105 bp *Hae*III DNA fragment in pFR6 contains the sequence CAAATTG (IS5: 479–485) which qualifies as a Pribnow box and, in the –35 region, the sequence CTTGGTAGCC^{29,30}. We suggest that this promoter, which we have called p_{5s} , is the transcription start signal for the IS5 small gene. The position of the p_{5s} promoter is thus different from its localization on the basis of sequence analysis (ref. 20). There are two possible translation start codons for the IS5 small reading frame at IS5: 495 and IS5: 525, coding for proteins of 118 and 108 amino acids respectively. The location of the Pribnow box at IS5: 479–485 suggests that translation starts at IS5: 525, where a translation start codon is preceded by a Shine–Dalgarno³¹ sequence. The absence of activity of p_{5s} *in vivo* may indicate that its activity is negatively regulated, poss-

ibly by the other IS5 elements present in the host strain (see below).

The positions of both promoters have been confirmed by the electron microscopic analysis of transcription complexes with purified *E. coli* RNA polymerase (P. Langridge, personal communication).

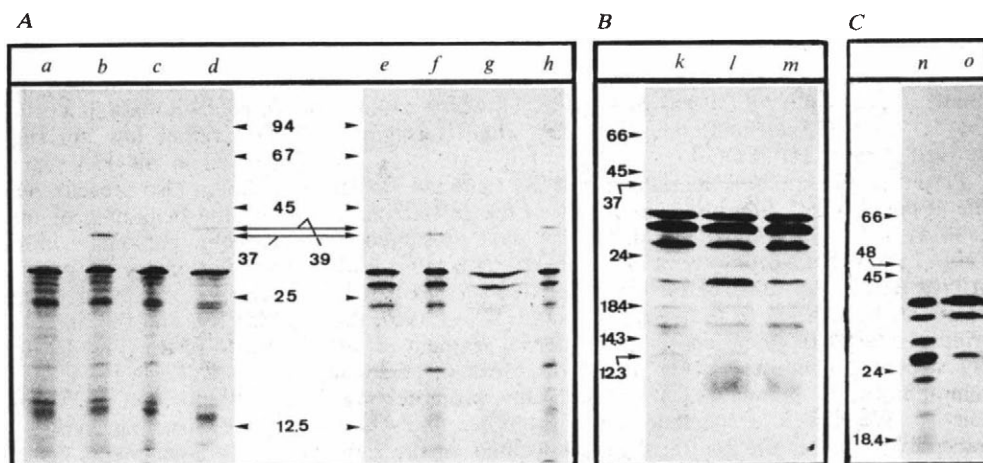
Expression of two proteins from opposite strands of IS5 DNA

Plasmids containing all or part of IS5 DNA (Fig. 1) were used to direct the incorporation of ^{35}S -methionine into proteins synthesized in a cell-free protein-synthesizing system and in minicells (Fig. 2A). In both systems, pFR1, which contains the entire IS5 element, directed the synthesis of a protein with an apparent molecular weight of 37,000 (Fig. 2A, lanes b, f). This protein was not detectable when the parental vector DNA (pML2) was used as template (lanes a, e). Another plasmid (pML3), containing only the long *Eco*RI arm of IS5 (IS5: 1–1,097), directed the synthesis of the pML2 vector proteins only (Fig. 2A, lanes c, g). This is consistent with the analysis of IS5 sequence data, which leads to the prediction that the NH₂-terminus and promoter of the IS5 large gene are located on the small *Eco*RI arm of IS5 (IS5: 1,097–1,195) with direction of transcription from right to left.

Correlation between the IS5 large gene of 326 codons and the 37,000 molecular weight (MW) protein is further supported by the data obtained using pHL128 (ref. 32) as template. In pHL128, a large IS5 fragment from the intact right end to the penultimate codon of the large gene at position IS5: 158 on the left side has, by *Bgl*II–*Bam*HI fusion, been linked to a foreign DNA fragment containing the *lacUV5* promoter sequence. This fusion eliminates the carboxy terminus of the IS5 protein and the reading frame is extended into the *lac* promoter fragment. The resulting fusion protein should contain 344 amino acids. As predicted, pHL128 directed the synthesis of a new protein with an apparent molecular weight of 39,000, about 2,000 larger than the original IS5 protein (Fig. 2A, lanes d, h).

When ^3H -leucine, the most abundant amino acid in the IS5 small gene coding sequence (18/108 amino acids^{20–22}), was used for protein labelling using pFR1 and pML3 as templates, a second protein with an apparent molecular weight 12,300 was detected. This protein was not observed with the parental vector pML2 as template (Fig. 2B). The λ DNA sequences present in pFR1 and pML3, as well as the different vector-insert fusion sequences in these hybrid plasmids, do not contain any reading frame long enough to account for the 12,300-MW protein (G.

Fig. 2 Expression of proteins encoded by IS5 in a cell-free protein-synthesizing system and in minicells. **A**, results obtained with ^{35}S -methionine. Cell-free protein synthesis was carried out as described in Table 1 legend with ^{35}S -methionine (1,007 Ci mmol⁻¹; 20 μCi per 100 μl assay). Minicell-producing strain P678-54 was transformed with the different plasmid DNAs, inoculated 1:50 into LB medium containing ampicillin (30 $\mu\text{g ml}^{-1}$) and grown overnight. Isolation and labelling of minicells were performed as described previously²⁴ using 0.5 A_{600} units of minicells and 40 μCi ^{35}S -methionine (1,007 Ci mmol⁻¹) in a total volume of 0.5 ml. Proteins (5×10^5 c.p.m. per track) were separated on 12.5% SDS-polyacrylamide gels⁵¹ and the gels were dried and autoradiographed. The position of standard protein size markers are indicated ($\times 10^3$). Lanes a–d: cell-free synthesis; lanes e–h: minicell synthesis. a, e, pML2; b, f, pFR1; c, g, pML3; d, h, pHL128. **B**, minicells were prepared as above and labelled with ^3H -leucine (65 Ci mmol⁻¹, 130 μCi per assay). Proteins (4×10^4 c.p.m. per track) were separated on 17% SDS-polyacrylamide gels which were fluorographed at –70 °C (ref. 52). Lane k, pML3; l, pFR1; m, pML2. **C**, minicells were prepared and assayed as described above with 110 μCi ^3H -leucine (65 Ci mmol⁻¹) and 110 μCi ^3H -arginine (53 Ci mmol⁻¹). Proteins (8×10^4 c.p.m. per track) were fractionated on 12.5% SDS-polyacrylamide gels and the radioactive bands visualized by fluorography. Lane n, pBR322; o, pFR11. The high background (seen previously²⁴) is due to the low specific activity of the tritiated amino acids and the necessity to optimize the minicells' protein-synthesizing capacity.



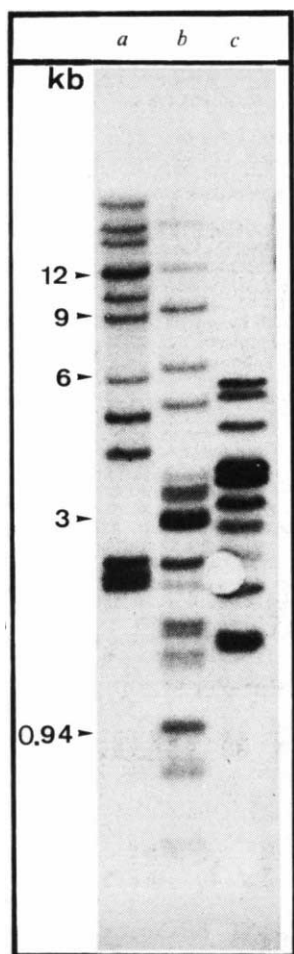


Fig. 3 Southern analysis of IS5 sequences in *E. coli* K-12 strain 431 Δ (*chlD-nadA*). *E. coli* DNA (5 μ g) was digested with restriction endonucleases *Bgl*II (a), *Pvu*II (b), *Hind*II (c) and electrophoresed on a 1% horizontal agarose gel. The fractionated DNA fragments were transferred to nitrocellulose by the method of Southern³⁷ and hybridized to the radioactively labelled IS5 internal *Bgl*II-*Eco*RI fragment. Note that *Pvu*II cuts once within IS5, between the *Bgl*II and the *Eco*RI site.

Hobom and M. Kröger, personal communication) and we conclude that this protein is the product of the IS5 small gene. The 12,300-MW protein was produced in the presence and absence of expression of the IS5 large gene (Fig. 2B, lanes k, l) and is, therefore, not a degradation product of the 37,000-MW protein.

Further support for expression of the IS5 small gene in minicells was obtained by gene fusion studies. In plasmid pFR11 (Fig. 1), the left arm of IS5 DNA up to the fourth codon from the end of the small gene was fused, in frame, to the coding sequence of the large tetracycline-resistance gene of pBR322. The resulting fusion protein would be expected to have a molecular weight of 48,800 if the IS5 small protein translational start is used. In accordance with this prediction, a 48,000-MW protein was encoded by pFR11 in minicells provided with tritiated leucine and arginine (which comprise 118 (59+59) out of 450 amino acids of the fusion protein) (Fig. 2C).

IS5 is present in multiple copies in *E. coli* K-12 genome

Most of the insertion elements found in the *E. coli* genome seem to be present in multiple copies (IS1³³, IS2³⁴ and IS3³⁵), with the exception of IS4 of which only a single copy has been found in most *E. coli* strains tested³⁶. Hybridization data using the gel blotting technique described by Southern³⁷ revealed 10–12 copies of IS5 in the DNA of the *E. coli* K-12 strain used (Fig. 3). This strain is a derivative of W8, a prototrophic *E. coli* K-12 from J. Weigle's collection. Interestingly, the same number of IS5 copies has recently been determined for four different *E. coli* K-12 strains kept separately for many years^{21,53}.

Discussion

Our data on the functional analysis of IS5 are summarized in Fig. 4.

From the DNA sequence of IS5, two overlapping genes can be proposed: a small one of 324 bp which is completely contained within a large gene of 978 bp. These genes have identical codon registers although transcription would have to proceed in opposite directions. They are preceded by signal structures for the initiation of RNA and protein synthesis. Our data present evidence that the transcription signal structures are indeed functional: RNA synthesis is initiated from the cloned promoter of the large gene both *in vivo* and in a cell-free system. Both genes are transcribed and translated in minicells and the gene products identified correspond in molecular weight to the values expected from the coding regions of their genes (Fig. 2). Activity of the isolated promoter *p*_{5s} on the other hand was only observed in the cell-free system and not *in vivo*. If we assume that both IS5 genes carry functions required for transposition, this observation may indicate a complex regulation of the small gene and, hence, of transposition on the level of transcription. Indeed, regulation of transposition has been observed for transposons Tn5³⁸ and Tn10³⁹ which are flanked by IS elements as well as for Tn3^{9–11,40}. The data presented here are consistent with the hypothesis that the small gene may require its own product *in vivo* for functioning and/or it underlies negative control mechanisms.

The data provide evidence for the unprecedented case of complete genes being transcribed and translated in an antiparallel fashion from the same nucleotide sequence. The highest degree of freedom for the sequence of coding triplets in two genes evolved in an antiparallel fashion from the same DNA fragment could be achieved by codon/codon register. DNA sequence analysis shows that the two IS5 genes are arranged in this mode.

Comparison of the available DNA sequence data for seven insertion elements revealed similar sequence organization with overlapping, antiparallel, long and short reading frames in codon/codon register in all^{8,17–23} but one⁴¹ case. This arrangement of genetic information may thus possibly be a common feature of IS elements.

We do not understand the selective advantage for a genetic structure with such a highly condensed information content, as shown by IS5. Selective pressure may have been favourable for a small size of genetic elements which themselves lack selective genes and face the elimination mechanism of the bacterial host cell. Alternatively, selection for a limited size of the element may lie in some step of the yet unknown mechanism by which insertion elements transpose.

We thank K. McKenney for providing plasmid pKO1 before publication, W. Szybalski for bacteriophage λ KH100, M. Kröger and G. Hobom for communicating and discussing results before publication, and P. Langridge and C. F. Beck for carefully reading the manuscript. This work was supported by grant Ra276/2 from the Deutsche Forschungsgemeinschaft.

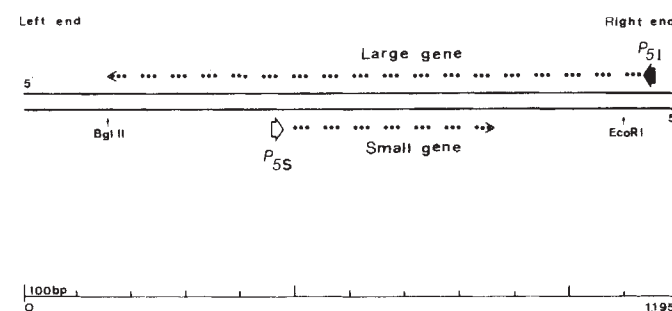


Fig. 4 Schematic presentation of the functional structures on IS5. *P*₅₁ and *P*_{5s} denote the locations of the promoters functionally detected on IS5, preceding the IS5 large gene and IS5 small gene respectively. The large arrows indicate the direction of transcription. Dots represent the codon/codon register of the triplets used by the two IS5 genes.

Received 15 February; accepted 18 March 1982.

1. Starlinger, P. *Plasmid* **3**, 241–259 (1980).
2. Calos, M. P. & Miller, J. H. *Cell* **20**, 579–595 (1980).
3. Kleckner, N. A. *Rev. Genet.* **15**, 341–404 (1981).
4. Grindley, N. D. F. & Sherratt, D. *Cold Spring Harb. Symp. quant. Biol.* **43**, 1257–1262 (1978).
5. Shapiro, J. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1933–1937 (1979).
6. Arthur, A. & Sherratt, D. *Molec. gen. Genet.* **175**, 267–274 (1979).
7. Ohtsubo, E., Zenilman, M. & Ohtsubo, H. *Proc. natn. Acad. Sci. U.S.A.* **77**, 750–754 (1980).
8. Grindley, N. D. F. & Joyce, C. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7176–7180 (1980).
9. Heffron, F., McCarthy, B. J., Ohtsubo, H. & Ohtsubo, E. *Cell* **18**, 1153–1163 (1979).
10. Gill, R., Heffron, F., Dougan, G. & Falkow, S. *J. Bact.* **136**, 742–756 (1978).
11. Heffron, F., So, M. & McCarthy, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6012–6016 (1978).
12. Altenbuchner, J., Choi, C. L., Grinsted, J., Schmitt, R. & Richmond, M. H. *Genet. Res.* **37**, 285–289 (1981).
13. Schmitt, R., Altenbuchner, J. & Grinsted, J. in *Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids* (ed. Levy, S. B.) 359–370 (Plenum, New York, 1981).
14. Harshey, R. M. & Bukhari, A. I. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1090–1094 (1981).
15. Jordan, E., Saedler, H. & Starlinger, P. *Molec. gen. Genet.* **102**, 353–363 (1968).
16. Shapiro, J. A. *J. molec. Biol.* **40**, 93–105 (1969).
17. Johnsrud, L. *Molec. gen. Genet.* **169**, 213–218 (1979).
18. Ghosal, D., Sommer, H. & Saedler, H. *Nucleic Acids Res.* **6**, 1111–1122 (1979).
19. Klaer, R., Kühn, S., Tillmann, E., Fritz, H. J. & Starlinger, P. *Molec. gen. Genet.* **181**, 169–175 (1981).
20. Kröger, M. & Hobom, G. *Nature* **297**, 159–162 (1982).
21. Schoner, B. & Kahn, M. *Gene* **14**, 165–174 (1981).
22. Engler, J. A. & Van Bree, M. P. *Gene* **14**, 155–163 (1981).
23. Bernardi, A. & Bernardi, F. *Nucleic Acids Res.* **9**, 2905–2911 (1981).
24. Trinks, K., Habermann, P., Beyreuther, K., Starlinger, P. & Ehring, R. *Molec. gen. Genet.* **182**, 183–188 (1981).
25. McKenney, K. et al. in *Gene Amplification and Analysis, Vol. II: Structural Analysis of Nucleic Acids* (eds Chirikjian, J. G. & Papas, T. S.) 383–415 (Elsevier, New York, 1981).
26. Blattner, F. R., Fiant, M., Hass, K. K., Twose, P. A. & Szybalski, W. *Virology* **62**, 458–471 (1974).
27. Szybalski, W. in *Regulatory Biology* (eds Copeland, J. C. & Marzluff, G. A.) 3–45 (Ohio State University Press, 1977).
28. Jones, M. O., Fischer, R., Herskowitz, I. & Echols, H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 150–154 (1979).
29. Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319–354 (1979).
30. Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269–281 (1980).
31. Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1342–1346 (1974).
32. Lusk, M., Kröger, M. & Hobom, G. *Cold Spring Harb. Symp. quant. Biol.* **45**, 173–176 (1980).
33. Nyman, K., Nakamura, K., Ohtsubo, H. & Ohtsubo, E. *Nature* **289**, 609–612 (1981).
34. Saedler, H. & Heiss, B. *Molec. gen. Genet.* **122**, 267–277 (1973).
35. Deonier, R., Hadley, R. & Hu, S. J. *Bact.* **137**, 1421–1424 (1979).
36. Klaer, R. et al. *Cold Spring Harb. Symp. quant. Biol.* **45**, 215–224 (1980).
37. Southern, E. J. *molec. Biol.* **98**, 503–517 (1975).
38. Biek, D. & Roth, J. R. *Cold Spring Harb. Symp. quant. Biol.* **45**, 189–191 (1980).
39. Beck, C. F., Moyed, H. & Ingraham, J. L. *Molec. gen. Genet.* **179**, 453–455 (1980).
40. Chou, J., Casadaban, M. J., Lemaux, P. & Cohen, S. N. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4020–4024 (1979).
41. Auerwald, E. A., Ludwig, G. & Schaller, H. *Cold Spring Harb. Symp. quant. Biol.* **45**, 107–113 (1980).
42. Zubay, G., Chambers, D. A. & Cheong, L. C. in *The Lactose Operon* (eds Beckwith, J. R. & Zipser, D.) 375–391 (Cold Spring Harbor Laboratory, New York, 1970).
43. Wetekamp, W., Staack, K. & Ehring, R. *Molec. gen. Genet.* **116**, 258–276 (1972).
44. Wilson, D. B. & Hogness, D. S. *Meth. Enzym.* **8**, 229–240 (1966).
45. Stürber, D. & Bujard, H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 167–171 (1981).
46. Sutcliffe, J. G. *Cold Spring Harb. Symp. quant. Biol.* **43**, 77–90 (1978).
47. Lusk, M. & Botchan, M. *Nature* **293**, 79–81 (1981).
48. Wienand, U. & Feix, G. *FEBS Lett.* **98**, 319–323 (1979).
49. Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).
50. Birnboim, H. C. & Doly, J. *Nucleic Acids Res.* **7**, 1513–1523 (1979).
51. Laemmli, U. K. *Nature* **227**, 680–685 (1970).
52. Bonner, W. M. & Laskey, R. A. *Eur. J. Biochem.* **46**, 83–88 (1974).
53. Nakamura, K. & Inouye, M. *Molec. gen. Genet.* **183**, 107–114 (1981).

Controlled transcription of a human α -interferon gene introduced into mouse L cells

Ned Mantei & Charles Weissmann

Institut für Molekularbiologie I, Universität Zürich, Hönggerberg, 8093 Zürich, Switzerland

Mouse L cells were transformed with a cloned chromosomal human interferon- α_1 gene, linked to a herpes thymidine kinase gene as selectable marker. In 8 out of 12 cell lines containing the human interferon gene, infection with Newcastle disease virus stimulated synthesis of messenger RNA for the human interferon—in parallel with that of the endogenous mouse interferon genes.

HUMAN interferons are assigned to three major classes¹: α , β and γ . The α -interferons (IFN- α) are encoded by a gene family consisting of a dozen or more non-allelic members^{2–9}. IFN- α preparations derived from induced human leukocytes or lymphoblastoid cell lines consist of five or more^{10,11} species, indicating that many of the genes are expressed *in vivo*.

Normal cells usually do not produce interferon and do not contain detectable levels of interferon mRNA¹². Treatment with inducers such as viruses or double-stranded RNA leads to the transient formation of IFN- α and/or IFN- β mRNA as well as interferon, 4–24 h after induction¹.

We show here that mouse L cells transformed with a cloned chromosomal IFN- α_1 gene produce correctly initiated IFN- α_1 mRNA in parallel with the appearance of mouse interferon mRNA after induction with Newcastle disease virus (NDV), but not in normal growth conditions. Thus the foreign gene seems subject to the normal control mechanisms of the host cell. A preliminary report of this work has been presented at the International Meeting on the Biology of the Interferon System, Rotterdam (21–24 April 1981).

Human IFN- α transcripts with correct 5' termini are produced in virus-induced transformed mouse cells

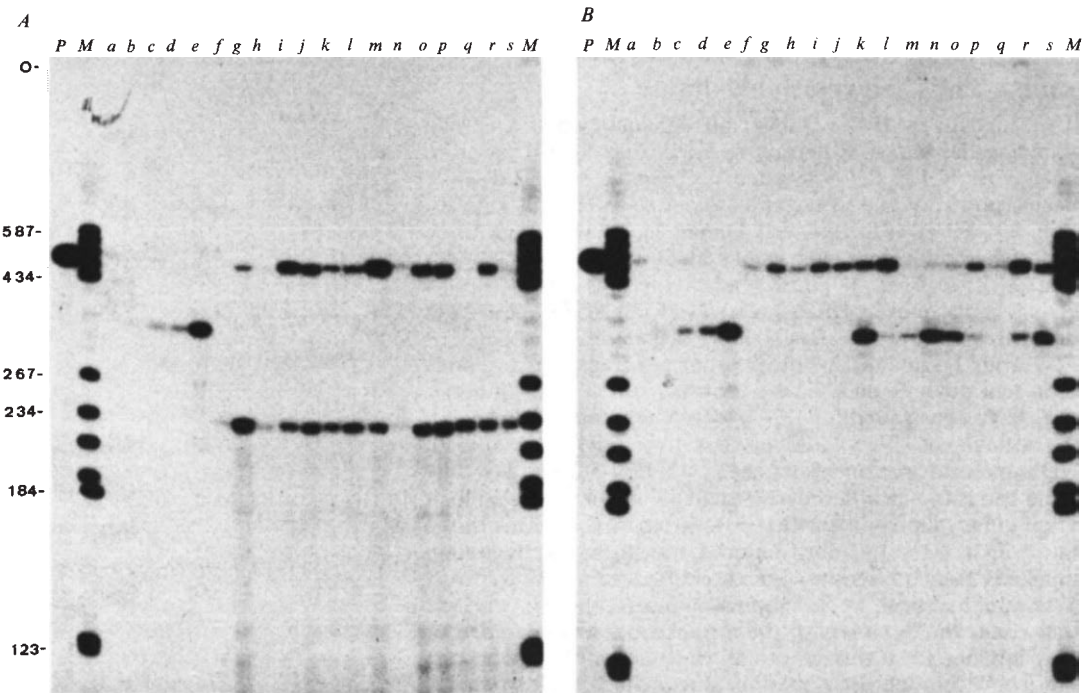
The hybrid phage ML-Ch (*Eco*)/Hchr(AH)IFN-35 carries the gene for human IFN- α_1 (ref. 3). A 7.5-kilobase (kb) *Bam*HI

fragment from this phage DNA, containing the IFN- α_1 gene, 5.4 kb of 5'-flanking and 1.2 kb of 3'-flanking sequences, was subcloned into the *Bam*HI site of pBR322 (pChr35-BB; S. Nagata, unpublished). The plasmid TK-M2 carries a thymidine kinase (TK) gene from herpes simplex virus type I (ref. 13). The two plasmids were cleaved at their single *Sal*I sites and joined by *T*₄ DNA ligase to form concatamers. LMTK cells were transfected with the concatamers using the calcium phosphate method^{14,15}. After 21 days in hypoxanthine/aminopterin/thymidine medium¹⁶, TK⁺ colonies were picked and expanded.

The IFN- α_1 DNA content of 13 TK⁺ cell lines (M2/IFN-1 to M2/IFN-13) was estimated by Southern blotting of *Eco*RI-cleaved cell DNA^{17,18} and hybridization with ³²P-labelled nick-translated IFN- α_1 cDNA¹⁹. All cell lines of the M2/IFN series except M2/IFN-3 contained IFN- α_1 DNA, ~0.25 to more than 20 copies per haploid genome (data not shown).

The human IFN- α_1 mRNA content of the cellular RNA was measured by an S₁ mapping procedure²⁰. A 480-base pair (bp) *Mbo*II-*Bgl*II fragment of the human chromosomal DNA, which spans the 5' terminus of the mRNA (see Fig. 6 of ref. 3), was 5'-terminally labelled at the *Bgl*II end. The denatured fragment was hybridized to the RNA, digested with S₁ nuclease and analysed by polyacrylamide gel electrophoresis²¹. Correctly initiated RNA is predicted to protect a ³²P-labelled fragment of 330 nucleotides³. As little as 0.14 ng poly(A)⁺ RNA from interferon-producing human leukocytes¹⁹ gave detectable

Fig. 1 Detection of human IFN- α_1 RNA with correct 5' termini in transformed mouse L cells. TK-negative mouse L cells (LMTK⁻) were transfected with concatamers of the plasmids TK-M2 (ref. 13) and pChr35-BB, which contains a 7.5-kb *Bam*HI fragment from the IFN- α_1 chromosomal DNA. 1.6 μ g TK-M2 and 9 μ g pChr35-BB were cleaved with *Sal*I and joined with ligase. Ten 9-cm plates of LMTK⁻ were transfected and TK-positive colonies selected as described previously^{15,23}. Control cells were transfected with *Sal*I-cleaved TK-M2 without ligation. RNA was prepared as described in ref. 27. To prepare the probe for S₁ mapping, plasmid pChr35-HB α (ref. 3) was cleaved with *Bgl*II (New England Biolabs) and the 5' termini labelled with ³²P-phosphate¹². After *Mbo*II cleavage, the 480-bp fragment was isolated. Probe (0.01 pmol, 27,000 c.p.m.) was denatured, hybridized in 80% formamide to 50 μ g cell RNA, digested with S₁ nuclease (P-L Biochemicals) and analysed by polyacrylamide gel electrophoresis on a 5% gel, all as described earlier^{12,21}. A, RNA from uninfected cells. Lanes a–e, 0, 0.14, 0.42, 1.25 and 3.75 ng poly(A)⁺ RNA from interferon-producing human leukocytes¹⁹. Lane f, RNA from cell line M2/S-33, transformed only with TK-M2. Lanes g–s, RNA from cell lines M2/IFN-1 to M2/IFN-13, transformed with TK-M2 linked to pChr35-BB. B, analysis of RNA from transformed cells infected with NDV. Transformed cells were grown for three generations with hypoxanthine and thymidine but without aminopterin, then infected with an optimal dose of NDV. RNA was extracted after 11 h. Samples as in A. Lane P, undigested probe. Lane M, size marker: pBR322 digested with *Bsp*I and 5'-³²P-labelled. Exposure for 21 days at -70 °C with an Ilford intensifying screen.



amounts of this fragment (Fig. 1A, lane b), whereas 50 μ g RNA from transformed cell lines M2/IFN-1 to M2/IFN-13 (Fig. 1A, lanes g–s) did not. There are therefore fewer than 0.5 human IFN- α_1 mRNA molecules with correct 5' termini per mouse cell transformed with IFN- α_1 DNA (see below for calculations). The 480-nucleotide fragment corresponds to completely protected probe, and is due, as shown below, to protection by complementary RNA longer than the probe and/or to self-renaturation. RNA from all mouse lines tested, including the line M2/S-33, transformed only with TK-M2 (Fig. 1A, lane f), gave a labelled fragment apparently 230 nucleotides long. We later found that this fragment and one of ~200 nucleotides (see Fig. 2) appear on electrophoresis of large amounts of undigested probe or undigested probe mixed with nuclease S₁-digested cell RNA (data not shown).

To test whether gene expression could be induced, the transformed cell lines were infected with NDV and RNA prepared 11 h later. As shown by S₁ mapping, RNA from 8 of 13 lines gave the 330-nucleotide fragment (Fig. 1B) indicative of correctly initiated human IFN- α_1 mRNA. The level of this mRNA was estimated from the amount of cell RNA per cell (~25 pg)²², the molecular weight of IFN- α_1 mRNA (~330,000) and the intensities of the 330-nucleotide bands relative to those given by RNA from induced human leukocytes, ~0.4% of which is IFN- α_1 RNA (N.M., unpublished). We estimate that eight of the cell lines contained ~0.5–10 IFN- α_1 transcripts with correct 5' termini per cell. No transcripts were found in cell line M2/S-33, transformed with TK-M2 alone (Fig. 1B, lane f).

To examine the kinetics of IFN- α_1 induction, cultures of M2/IFN-5 and of M2/S-34 (transformed with TK-M2 alone) were infected with NDV or mock-infected. After 20 h, media from both infected cultures contained about 4,000 units ml⁻¹ of mouse interferon, indicating that mouse interferon was induced in both cell lines. RNA was prepared after 4, 11, 20 and 27 h, and poly(A)⁺ and poly(A)⁻ RNA separated by

chromatography on oligo(dT)-cellulose^{23,24}. IFN- α_1 RNA was assayed by the S₁ mapping procedure described above. In poly(A)⁺ RNA from infected cultures of M2/IFN-5, correctly initiated IFN- α_1 RNA was present after 11 h (Fig. 2, lane l) and, in lower amounts, after 20 h (lane c), but not after 4 h (lane a) or 27 h (lane e). Neither cell line M2/S-34 (lanes n, o) nor mock-infected M2/IFN-5 (lanes b, d, f, m) ever gave a 330-nucleotide fragment. The mobilities of the 330-nucleotide fragments given by M2/IFN-5 RNA (Fig. 2, lane l) and by human leukocyte RNA (lane k) are the same, implying that the 5' ends of the RNAs are identical. The signal from poly(A)⁺ RNA from 150 μ g M2/IFN-5 RNA at 11 h after infection (lane l) was about 15 times stronger than that from poly(A)⁺ RNA from 50 μ g of the same cell RNA (lane p). About 80% of the IFN- α_1 RNA is therefore polyadenylated.

The kinetics of induction of mouse interferon mRNA was followed by injecting samples of poly(A)⁺ cell RNA into *Xenopus* oocytes and determining antiviral activity on mouse cells. Activity was undetectable at 4 h, reached a maximum at 11 h and dropped to about one-tenth of maximum at 20 h after infection.

The identification of human IFN- α_1 mRNA in virus-induced mouse cells depends on the formation of perfect or almost perfect hybrids between the mRNA and the human DNA probe in stringent hybridization conditions. Human and mouse IFN- α coding sequences differ in about 26% of their residues (G. Shaw and N.M., unpublished), so that only very imperfect hybrids could be formed even in non-stringent hybridization conditions. In fact, no fragments characteristic for correctly initiated human IFN- α_1 mRNA were found when 5' probes were hybridized to RNA from interferon-producing mouse cells transformed with TK plasmid alone (Fig. 1B, lane f). The results of the S₁ mapping of RNA from virus-induced mouse cells containing the human IFN- α_1 gene therefore cannot be explained by cross-reaction with mouse interferon mRNA.

Human IFN- α transcripts with incorrect 5' termini in uninduced, transformed mouse cells are relatively stable

The 3' ends of the IFN- α_1 transcripts were mapped with a probe spanning the known 3' terminus. Even after hybridization to only yeast RNA, the S_1 -treated probe gave several discrete bands, probably due to incomplete digestion (Fig. 3, lane *a*). Poly(A)⁺ RNA from induced human leukocytes gave three additional protected fragments of 550, 600 and 780 nucleotides (Fig. 3, lanes *b*, *c*, *f*). The 3' end of the plus-strand IFN- α_1 cDNA in plasmid Z-pBR322 (*Pst*)/HcIF-2h is 554 nucleotides downstream from the *Bgl*II site¹², accounting for the 550-nucleotide fragment. The other protected fragments probably represent other 3' ends further downstream. Two or more 3' ends have been found for IFN- α_2 RNA⁷ and other RNAs^{25,26}. An artefact of the S_1 assay is less likely. RNA from both NDV-infected and mock-infected M2/IFN-5 cells gave the same bands found with induced human leukocyte RNA, albeit at different relative intensities (Fig. 3, lanes *d*, *e*). This shows that both induced and mock-induced, transformed cells contain human IFN- α_1 transcripts with correct 3' termini. As no correct 5' termini had been found in mock-induced cells, we conclude that either the 5' moiety of these transcripts was degraded or they initiated far upstream of the 'cap' site. In the latter case, the labelled probe used previously to map 5' ends would be

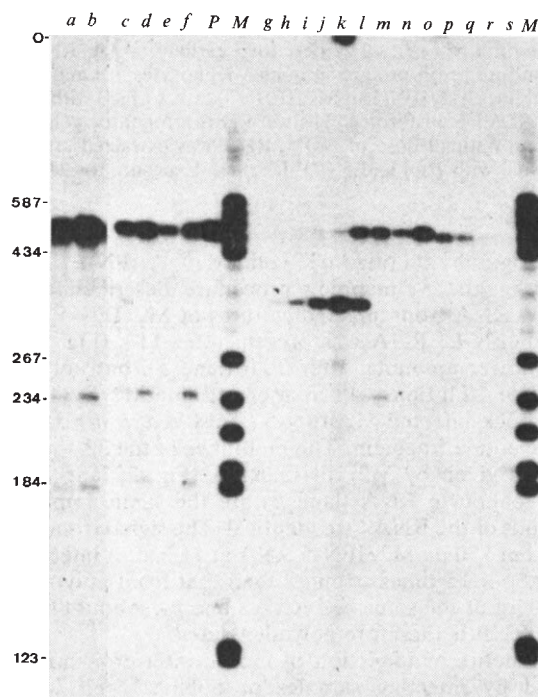


Fig. 2 5' termini of human IFN- α_1 RNA in transformed mouse cells at different times after infection with NDV. M2/IFN-5 and M2/S-34 cells (see Fig. 1 legend) were infected with NDV or were mock-infected. RNA was prepared at 4, 11, 20 and 27 h after the end of the adsorption period and fractionated by chromatography on oligo(dT)-cellulose^{23,24} (P-L Biochemicals, type 7). Poly(A)⁺ RNA derived from 150 μ g cell RNA and poly(A)⁻ RNA from 50 μ g cell RNA were analysed as for Fig. 1. Lanes *g-k*, 0, 0.16, 0.5, 1.5 and 4.5 ng poly(A)⁺ RNA from interferon-producing human leukocytes¹⁹. Lanes *a*, *c*, *e* and *i*, poly(A)⁺ RNA from M2/IFN-5 at 4, 20, 27 and 11 h after NDV infection. Lanes *b*, *d*, *f* and *m*, poly(A)⁺ RNA from M2/IFN-5 at 4, 20, 27 and 11 h after mock infection. Lanes *n*, *o*, poly(A)⁺ RNA prepared 11 h after infection (*n*) or mock infection (*o*) of M2/S-34. Lanes *p-s*, poly(A)⁻ RNA prepared 11 h after infection (*p*, *r*) or mock infection (*q*, *s*) of M2/IFN-5 (*p*, *q*) and M2/S-34 (*r*, *s*). Lanes *P* and *M* as in Fig. 1. The autoradiograph was exposed for 25 days.

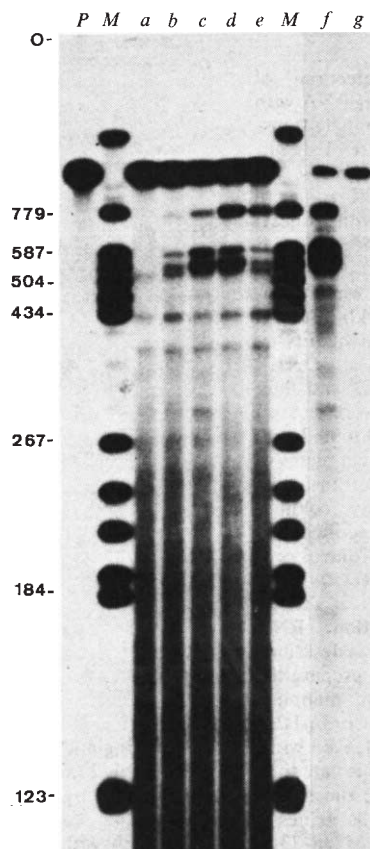


Fig. 3 S_1 mapping of 3' ends of human IFN- α_1 RNA. The plasmid pChr35-HB α containing the human chromosomal IFN- α_1 sequence was incubated with *Bgl*II and then with T_4 DNA polymerase (P-L Biochemicals) at 15 °C in the presence of 0.1 mM dCTP³⁵. After phenol extraction and ethanol precipitation, the termini were 'filled in' with the Klenow fragment of DNA polymerase (Boehringer) in the presence of 0.2 mM each dCTP, dGTP and TTP, and 2.4 μ M [α -³²P]dATP (NEN). Following cleavage with *Bsp*I, the resulting 1,000-bp *Bgl*II-*Bsp*I fragment (see map in ref. 3) was isolated. S_1 analysis as for Fig. 1. Lanes *a-c*, *f*, *g*, 0, 1.2, 3.6, 300 and 0 ng poly(A)⁺ RNA from interferon-producing human leukocytes¹⁹. Lanes *d*, *e*, poly(A)⁺ RNA (from the preparation described in Fig. 2 legend) derived from 120 μ g cell RNA of cell line M2/IFN-5, either 11 h after NDV-infection (*d*) or mock infection (*e*). Lanes *P* and *M* as in Fig. 1. Exposure for 25 days, except that lanes *f* and *g* were exposed for 16 h.

protected over its full length, giving a signal indistinguishable from renatured probe.

To determine whether IFN- α_1 RNA in mock-infected cells originated upstream of the cap site, we used a probe with a 5'-³²P-labelled *Eco*RI end at nucleotide 633 in the IFN- α_1 gene, extending upstream to nucleotide -675 in the 5'-flanking region and continuing 103 nucleotides into the adjoining pBR322 DNA (see Fig. 4). All '3' halves' of IFN- α_1 RNA found in the experiment of Fig. 3 will be detected by this probe, as the *Eco*RI site lies 372 nucleotides downstream from the labelled *Bgl*II end used for the 3'-end mapping. Transcripts initiated further upstream than position -675 will yield a protected fragment of ~1,375 nucleotides which is easily distinguished from undegraded probe (~1,480 nucleotides). Figure 4*B*, lane *e*, shows that RNA from NDV-infected M2/IFN-5 yielded as the major product the 700-nucleotide fragment expected for correctly initiated RNA, with lesser amounts of 1,100- and 1,375-nucleotide fragments. RNA from mock-infected M2/IFN-5 protected fragments of 1,100 and 1,375 nucleotides only (Fig. 4*B*, lane *f*). Mapping of 3' ends of RNA from the same batch showed that the 5' ends detected account for most or all of the 3' ends (data not shown).

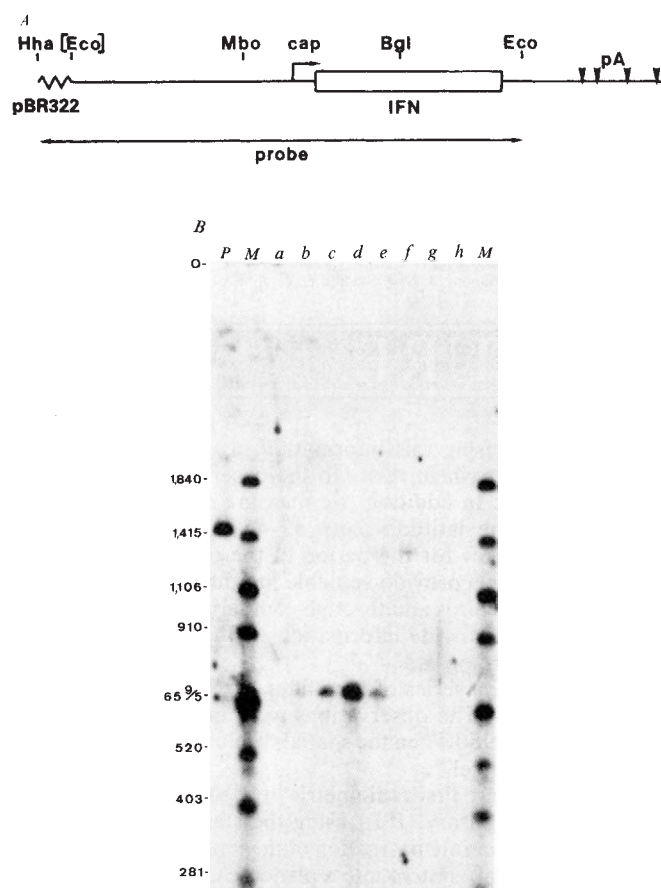


Fig. 4 S₁ mapping of 5' ends of human IFN- α_1 RNA in transformed mouse cells. M2/IFN-5 and M2/S-34 cells were infected with NDV or mock-infected, and poly(A)⁺ RNA prepared after 11 h. A, the plasmid pChr35-675 is a derivative of pChr35-HB α (ref. 3) in which the EcoRI fragment between -675 and the EcoRI site of pBR322 was deleted and the new EcoRI site destroyed [Eco]. pChr35-675 was cleaved with HhaI and the 5' termini labelled. After cleavage with HhaI and PstI, the 1,480-bp HhaI-EcoRI fragment was isolated. B, mapping of 5' ends. The fragment described above (0.01 pmol, 40,000 c.p.m.) was hybridized to poly(A)⁺ RNA derived from 125 μ g cell RNA, digested with S₁ nuclease at 24 °C and the products analysed by electrophoresis through a 1.6% alkaline agarose gel³⁶. Lanes a-d, 0, 0.42, 1.25 and 3.75 ng poly(A)⁺ RNA from interferon-producing human leukocytes. Lanes e, f, RNA from cell line M2/IFN-5 after infection (e) or mock infection (f). Lanes g, h, RNA from cell line M2/S-34 after infection (g) or mock infection (h). Lane P, undigested probe; lanes M, pBR322 cleaved with HincII and PvuII, plus pBR322 cleaved with AluI, both terminally labelled. Exposure for 14 days.

The S₁ mapping of poly(A)⁺ RNA from induced, transformed mouse cells, from the BglII site to the 5' and 3' termini and from the EcoRI site to the 5' termini, gives signals identical to those found with poly(A)⁺ RNA from induced human leukocytes. We conclude that the induced, transformed mouse cells contain complete human IFN- α_1 mRNA which has the correct termini and is polyadenylated.

Conclusions

Mouse L cells, transformed with a cloned human IFN- α_1 gene, produce a low level of incorrectly initiated interferon transcripts. Most, if not all, of these transcripts initiate upstream of the cap site. Transcripts initiated upstream of the cap site have also been found in the case of the rabbit β -globin gene and the chicken ovalbumin gene introduced into mouse L cells^{27,28}. When mouse L cells transformed with the IFN- α_1 gene are infected with NDV, correctly initiated, polyadenylated

IFN- α_1 transcripts appear, at levels of 0.5–10 strands per cell. We believe this to be due to *de novo* synthesis, and not to processing of the pre-existing overlong transcripts: if such processing were accidental, it is unlikely that newly formed 5' termini would coincide precisely with the 5' termini of natural IFN- α_1 mRNA. On the other hand, it is unlikely that, in physiological conditions, overlong transcripts give rise to mature interferon RNA by processing, because human leukocytes contain no IFN- α -related transcripts before induction¹². The time course of appearance and disappearance of the human IFN- α_1 RNA in mouse cells is similar to that of the mouse interferon mRNA in the same cells, suggesting a common control mechanism. Whereas mouse interferon was easily measured in the medium, human IFN- α_1 could not be detected, probably because the interferon mRNA level was low, and the specific antiviral activity of IFN- α_1 on human cells is lower by an order of magnitude than that of other human interferons²⁹.

Induction of interferon mRNA in normal cells could be explained in two ways: (1) although interferon mRNA is synthesized at a similar rate in non-induced and induced cells, transcripts are rapidly degraded in non-induced cells but stabilized in induced cells. (2) There is no (correct) transcription of interferon genes in non-induced cells, and induction activates transcription. The fact that non-induced transformed mouse cells contain about as many incorrectly initiated IFN- α_1 mRNA transcripts as induced cells suggests that, unless a correct 5' terminus is required for degradation, the stability of interferon transcripts is similar in both cases. Thus, induction would not suppress interferon mRNA degradation, but rather stimulate correct transcription of IFN- α genes, which does not occur to a measurable extent in non-induced cells.

The interferon gene flanked by 5.4 kb of 5'- and 1.2 kb of 3'-chromosomal DNA seems to contain sufficient information to respond to physiological induction, at least to some extent. We estimate that there is about 50 times less human than mouse interferon mRNA in an induced, transformed mouse cell line containing two IFN- α_1 genes per haploid genome; this may be due in part to the fact that the cells contain many more mouse interferon genes, probably more than 10 copies per haploid genome. In addition, there may be some incompatibility between the human and mouse system, and/or effects due to incorrect positioning of the transplanted genes in the mouse genome.

Recently, other cloned genes introduced into host cells have been found to respond to induction—the cloned mouse mammary tumour virus genome^{30,31}, the rat α_2 globulin gene³² and a *Drosophila* heat-shock gene³³. Thus, several systems now provide an approach to the identification of the DNA sequences responsible for control of transcription.

This work was supported by the Schweizerische Nationalfonds (3.481.79) and the Kanton of Zürich. We thank S. Nagata and U. Weidle for plasmids, C. Arnheiter for NDV, M. Streuli for injecting oocytes, P. Dierks and J. Fujisawa for [γ -³²P]ATP, and C. Arnheiter and J. Fujisawa for helpful advice.

Note added in proof: Ohno and Taniguchi³⁴ have found that mouse FM3A cells transformed with the human IFN- β gene produce IFN- β and increased amounts of IFN- β mRNA after treatment with NDV or poly(I) · poly(C). Similar results have been obtained by H. Hauser *et al.* (personal communication).

Received 31 December 1981; accepted 16 March 1982.

1. Stewart, W. E. II *The Interferon System* (Springer, New York, 1979).
2. Streuli, M., Nagata, S. & Weissmann, C. *Science* **209**, 1343–1347 (1980).
3. Nagata, S., Mantei, N. & Weissmann, C. *Nature* **287**, 401–408 (1980).
4. Goeddel, D. V. *et al. Nature* **290**, 20–25 (1981).
5. Nagata, S., Brack, C., Henco, K., Schamböck, A. & Weissmann, C. *J. Interferon Res.* **1**, 333–336 (1981).
6. Lawn, R. M. *et al. Science* **212**, 1159–1162 (1981).
7. Lawn, R. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5435–5439 (1981).
8. Owerbach, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3123–3127 (1981).
9. Brack, C., Nagata, S., Mantei, N. & Weissmann, C. *Gene* **15**, 379–394 (1981).
10. Allen, G. & Fantes, K. H. *Nature* **287**, 408–411 (1980).
11. Rubinstein, M. *et al. Archs Biochem. Biophys.* **210**, 307–318 (1981).
12. Mantei, N. *et al. Gene* **10**, 1–10 (1980).

13. Wilkie, N. M. *et al.* *Nucleic Acids Res.* **7**, 859–877 (1979).
14. Graham, & van der Eb, A. J. *Virology* **52**, 456–467 (1973).
15. Wigler, M., Pellicer, A., Silverstein, S. & Axel, R. *Cell* **14**, 725–731 (1978).
16. Szybalski, W., Szybalska, E. H. & Ragni, G. *Monogr. natn. Cancer Inst.* **7**, 75 (1962).
17. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
18. Wahl, G. M., Stern, M. & Stark, G. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3683–3687 (1979).
19. Nagata, S. *et al.* *Nature* **284**, 316–320 (1980).
20. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
21. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **7**, 1175–1193 (1979).
22. Parsons, J. T., Coffin, J. M., Haroz, R. K., Bromley, P. A. & Weissmann, C. *J. Virol.* **11**, 761–774 (1973).
23. Mantei, N., Boll, W. & Weissmann, C. *Nature* **281**, 40–46 (1979).
24. Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408–1412 (1972).
25. Tosi, M., Young, R. A., Hagenbüchle, O. & Schibler, U. *Nucleic Acids Res.* **9**, 2313–2323 (1981).
26. Nevins, J. R. & Darnell, J. E. *Cell* **15**, 1477–1493 (1978).
27. Dierks, P., van Ooyen, A., Mantei, N. & Weissmann, C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1411–1415 (1981).
28. Breathnach, R., Mantei, N. & Chambon, P. *Proc. natn. Acad. Sci. U.S.A.* **77**, 740–744 (1980).
29. Streuli, M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **78**, 2848–2852 (1981).
30. Hynes, N. E., Kennedy, N., Rahmsdorf, U. & Groner, B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2038–2042 (1981).
31. Buetti, E. & Diggelmann, H. *Cell* **23**, 335–345 (1981).
32. Kurtz, D. T. *Nature* **291**, 629–631 (1981).
33. Corces, V., Pellicer, A., Axel, R. & Meselson, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7038–7042 (1981).
34. Ohno, T. & Taniguchi, T. *Nucleic Acids Res.* **10**, 967 (1982).
35. Challberg, M. D. & Englund, P. T. *Meth. Enzym.* **65**, 39–45 (1980).
36. McDonnell, M. W., Simon, M. N. & Studier, F. W. *J. molec. Biol.* **110**, 119–146 (1977).

LETTERS TO NATURE

Dynamical features in the northern hemisphere of Saturn from Voyager 1 images

Garry E. Hunt, David Godfrey, Jan-Peter Müller & Robert F. T. Barrey

Laboratory for Planetary Atmospheres, Department of Physics and Astronomy, University College London, London WC1E 6BT, UK

Results of the analysis of Voyager measurements of divergence and inferred vertical velocity of convective cloud features in the northern hemisphere of Saturn are presented here. The winds in this region have been measured by tracking the cloud elements observed in sequences of images. The derived zonal winds imply that these convective features reside at the maximum of strong easterly jets centred at 41° N planetocentric latitude and that the jet is barotropically unstable. The divergences and local dynamics of the Saturn features are found to differ from those of the jovian equatorial plumes.

During the encounter of Saturn by Voyager 1, Smith *et al.*¹ reported observations of distinct cloud features occurring in the northern hemisphere of the planet's atmosphere. In the region bounded by 35–42° N, several bright V-shaped features have been seen, while at ~29° N, a distinct cloud oval is apparent (Fig. 1, feature 3.) (All latitudes in this article are planetocentric.) The large spot is particularly evident in the UV images and was seen to persist throughout the Voyager 1 observational period from 24 August to 10 November 1980. The V-shaped features varied greatly in shape during the observation period but persisted in their longitudinal distribution as they drifted in an easterly jet centred at 41° N. The changes in the visible appearance of these features suggests that they are convective clouds where the rapid variations in the shape and structure of these cloud systems is associated with local vertical motions. There is no evidence from the IR observation of Hanel *et al.*² that these features affect the temperature structure of the upper troposphere. However, the IRIS field of view is large compared with the visible cloud structures observed by the imaging system³. The IRIS instrument essentially measures the average temperature for the total region seen in the imaging frame. Indeed, any upper atmosphere aerosol layers present would further reduce any possible temperature contrasts.

In the atmosphere of Jupiter, we have shown that the equatorial plumes are convective cloud structures^{3,4} and their meteorological characteristics have been estimated by procedures developed using the University College London (UCL) Interactive Planetary Image Processing System⁵. We apply these techniques for the analysis of the Saturn observations and present here the first measurements of divergence of saturnian

cloud systems. Using this information, we then calculate the vertical velocities which relate to the observed changes in the cloud structures. In addition, we measure the wind speeds of the clouds in the latitude band 30–45° N and estimate the stability of the flow for the period of the observations. These new measurements provide valuable insight into the variations of the cloud structures and through comparison with our jovian studies⁴ this will provide information on the differences in the observed meteorologies.

For this study, a series of blue filter narrow angle images of Saturn was used. The observations were made for the period 3–8 November 1980 when the spatial resolution varied between 120 and 60 km pixel⁻¹.

The images were first radiometrically decalibrated at the Jet Propulsion Laboratory (JPL) using the standard procedures⁵. To provide an accurate navigation of the images that transforms the line and sample system into a planetocentric latitude/longitude coordinate system, two independent procedures were used. At JPL, the method used simultaneous wide and narrow angle frames to locate the centre of the planet from which the subsequent transformation is then determined⁵. An interactive procedure, developed at UCL, uses a 'least-squares' minimization method to fit an ellipse to the image, constrained by the spacecraft navigation and camera data. We have shown elsewhere⁶ that both of these methods produce similar results although the error in determining the position of the centre of the planet becomes greater than 1 pixel when 20% or less of the planet's limb is available (see ref. 7).

Measurements of velocities of the upper clouds in the Saturn atmosphere have been made by interactively tracking individual features observed in the images. The techniques have been shown⁶ to be accurate to within $\pm 3 \text{ ms}^{-1}$ in zonal and meridional winds at a spatial resolution of 60 km pixel⁻¹ and produce results that compare favourably with other schemes¹. The velocities are measured relative to the SYSTEM III radio period of 10 h 39.4 min (ref. 8) using images separated by time intervals of 10–20 h. Considerable contrast enhancement of the image was first necessary to show clearly the cloud elements to be tracked.

The measurements of the divergence of the V-shaped clouds were made by the procedures set out in ref. 4. The sequence of 10 blue narrow angle images were first map projected into a Mercator format centred at 41° N using a single pixel interpolation scheme developed at UCL⁶. Then, the divergence of these features was estimated from the changing cloud areas. We estimate the area of the convectively active region from the assumption that the brightness in the cloud system, which we have previously shown^{4,9,10} to be a reliable approach.

If we now assume that the vertical motion in the layer of outflow takes place over a vertical scale D , then the vertical velocity required to produce the observed divergence field is:

$$W \leq D \operatorname{div} \mathbf{V} \quad (1)$$

We estimate D as the scale height of the atmosphere at the cloud top, which will then enable an estimate to be made of the vertical motions associated with these cloud systems.

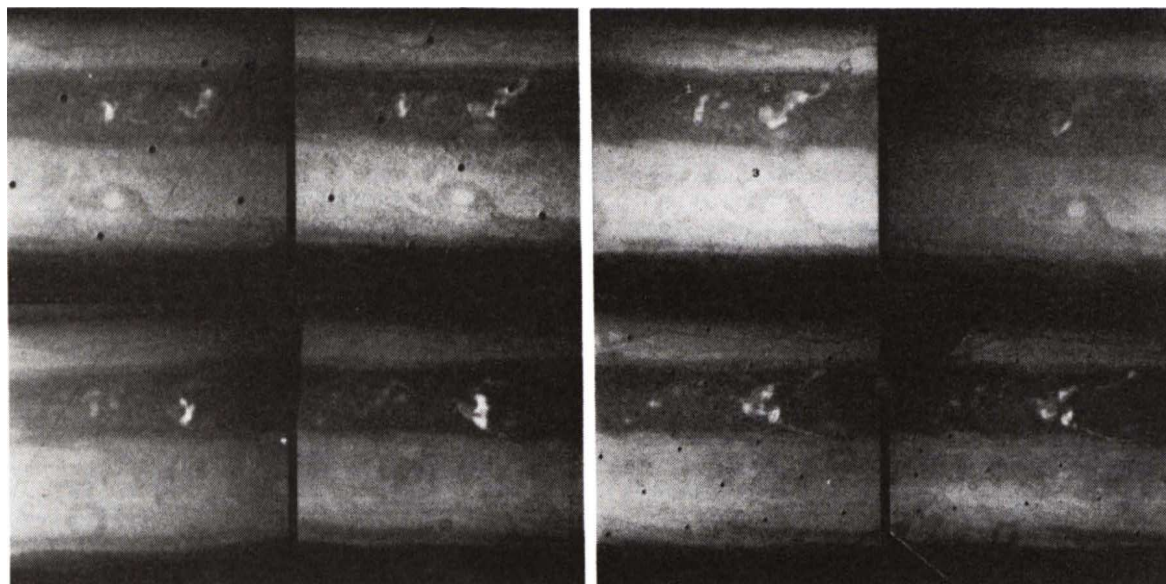


Fig. 1 Sequence of eight blue Mercator narrow angle images of the mid-latitudes of Saturn. The dates and times of the images are: 15:05:22, 3 November 1980; 11:57:22, 4 November 1980; 22:11:46, 4 November 1980; 08:25:22, 5 November 1980, 20:46:58, 5 November 1980; 17:22:10, 6 November 1980; 12:08:34, 8 November 1980; 22:24:34, 8 November 1980.

Figure 1 shows the variation in the V-shaped cloud structures for the period 15:05:22, 3 November to 22:24:34, 8 November 1980. All times are GMT at the spacecraft. This time sequence clearly indicates the large morphological change in the convective cloud structures. At the edge of feature 2, considerable developments in the clouds become apparent in the final three images of the sequence. Figure 1 shows less noticeable changes during this time.

The third main cloud feature seen in Fig. 1 is the UV spot. The prominence of this feature at short wavelengths is consistent with it being higher than its surrounding clouds. A well defined flow pattern can be seen to the west of this feature in Fig. 1, which is morphologically similar to the 'turbulent' regions to the west of the Great Red Spot and White Ovals on Jupiter.

In Fig. 2, we summarize the basic characteristics of the motions in the mid-latitude region bounded by 30–50° N. The mean zonal velocity is shown in Fig. 2a; a strong easterly jet is situated at 41° N which corresponds to the apex of the V features. We find also that the strongest shear occurs along the southern edge of this jet. The small inflection in the zonal velocity profile at 42–43° N is associated with small changes in the cloud morphologies.

To discuss the significance of these results and the stability of the flow, we estimate $\bar{u}$, $d^2\bar{u}/dy^2$, and their associated errors, using the procedures set out in ref. 11. The measured zonal flow is composed of contributions from the mean flow, the eddies and errors which arise from effects of spacecraft navigation, preprocessing of the data and uncertainties in the identification of individual features. Following the procedures of Ingersoll *et al.*¹¹, we estimate the error in $\bar{u}$, $\sigma(\bar{u})$ as $\pm 5 \text{ m s}^{-1}$. We compute $d^2\bar{u}/dy^2$ at latitudinal position y_n by:

$$(d^2\bar{u}/dy^2)_n = (\bar{u}_{n+2} - 2\bar{u}_n + \bar{u}_{n-2})/\Delta y_n^2$$

where $\Delta y_n = y_{n+1} - y_{n-1}$.

Then the error in $d^2\bar{u}/dy^2$ is $\sim \pm \sqrt{6}\sigma(\bar{u})/\Delta y^2$ (ref. 11).

The stability of the retrograde jet may be investigated by examining the barotropic stability condition for a zonal velocity $\bar{u}$ (see, for example, ref. 12), given by:

$$\beta - d^2\bar{u}/dy^2 \geq 0$$

where $\beta = (2\Omega/r) \cdot \cos \theta$, Ω is the planetary rotation, r the planetary radius (local radius or curvature), θ the planetodetic latitude and y the northerly direction. In Fig. 2b, we have superimposed the value of β with $d^2\bar{u}/dy^2$, while the additional

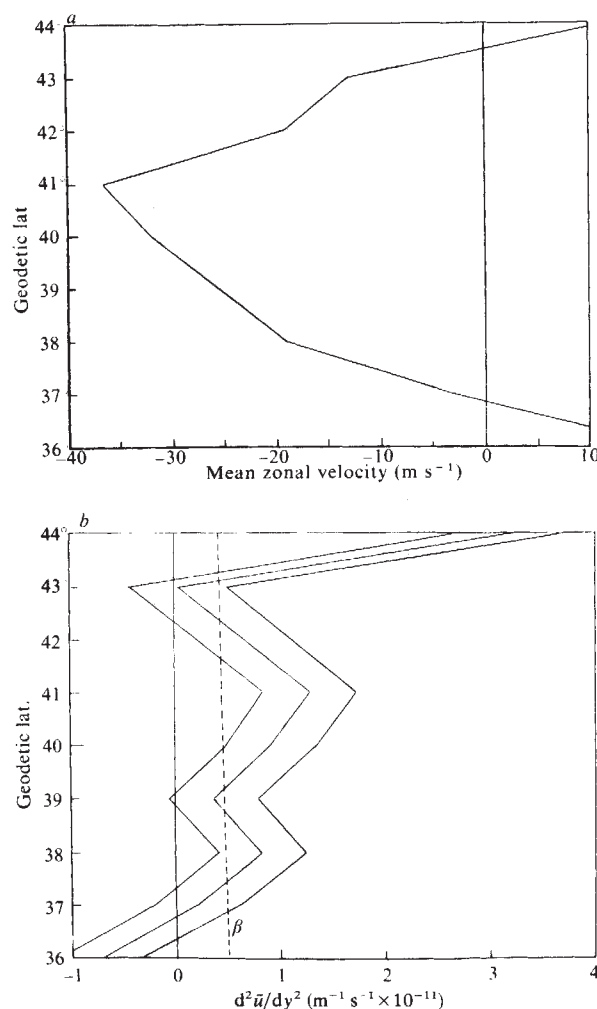


Fig. 2 The dynamical properties of the region 30–50° N derived in this study. *a*, The zonal wind $\bar{u}$ is shown with 1σ r.m.s. values. *b*, The curve for $d^2\bar{u}/dy^2$, together with the value of β at 41° N. The central curve is the variation of $d^2\bar{u}/dy^2$. The curves at the left and right show the variation in $d^2\bar{u}/dy^2$ when the errors are taken into account.

curves illustrate the errors in $d^2\bar{u}/dy^2$ estimated by the procedures set out above. We see that the flow between 40 and 41.5° N is barotropically unstable in more stringent conditions. The figure further suggests that the region extending to 37° N is close to being unstable in these conditions. The V-cloud features reside in the centre of the unstable region, which is also the maximum in the easterly jet. The stable region of the flow lies outside the latitudes of 37–42° N.

Figure 3 shows the divergence of these active regions. There is little change in the characteristics of the UV spot, which is apparent from the image sequence seen in Fig. 1. The changes in area shown in Fig. 2 for the convective features were used to determine the 90% of maximum brightness threshold contour using the procedures described in ref. 4. These variations in area of the active clouds may then be used to estimate the divergence of the features which are shown in Fig. 3. As expected, feature 2 is the most active, with values reaching $\sim 5 \times 10^{-5} \text{ s}^{-1}$. The other convective feature is less active, with a maximum value of $3 \times 10^{-5} \text{ s}^{-1}$. A negative divergence implies the dissipation of the cloud feature.

Comparing the results with those of our study⁴ of the jovian equatorial plumes, we find that the divergences are much stronger for the saturnian features. The largest Saturn values are more than three times greater than those associated with the most active jovian plume. Assuming that the anvil expansion we are measuring takes place at $\sim 500 \text{ mbar level}^2$ over a depth of a density scale height of 40 km (ref. 13), then the divergence maxima imply vertical motions of $\sim 2 \text{ m s}^{-1}$.

On Jupiter, these large scale motions are propagated throughout the troposphere³. Although the saturnian vertical motions are apparently stronger, the IR observations do not suggest any large scale perturbations in the temperature structure in a horizontal or vertical direction. It is possible that the larger aerosol layers in the troposphere of Saturn provide sufficient opacity to diminish the temperature contrasts. Furthermore, the IRIS observations are essentially the average response for the narrow angle field of view, so that these small scale convective clouds may not produce a sufficiently large signature to be noticeable.

The present estimates of larger vertical motions in the atmosphere of Saturn compared with Jupiter are consistent with the differences in the zonal velocity and internal structure of the two planets. The internal heat source on Saturn supplies approximately twice as much energy to the atmosphere as solar heating². This situation, together with the smaller gravity, which then leads to a larger atmospheric scale height on Saturn, is consistent with the larger convective motions discussed above.

From these first estimates of divergence and vertical motions of cloud features observed in the northern hemisphere of Saturn we find that the associated flows of the convective clouds are stronger than the similar cloud features observed in the atmosphere of Jupiter. The vertical velocities we estimate for these

motions are of a similar magnitude to large scale terrestrial convective storms (see, for example, ref. 9), although on Saturn these storms occupy a considerably greater spatial scale.

We find that the Saturn convective cloud features are situated in the maximum of the easterly jet which corresponds to a barotropically unstable region of the atmosphere (Fig. 2). It is possible that these mesoscale cloud features are generated by large scale convergence in the deep cloud layers beneath the visible surface of the planet. These stronger motions are consistent with the differences in the structures of Saturn and Jupiter.

We thank Glenn Garneau, Eric Korsmo and Gary Yagi at JPL for assistance. This work is supported by the Science and Engineering Research Council.

Received 8 March; accepted 1 April 1982.

1. Smith, B. A. *et al. Science* **212**, 183–191 (1981).
2. Hanel, R. A. *et al. Science* **212**, 192–201 (1981).
3. Hunt, G. E., Conrath, B. J. & Pirraglia, J. A. *J. geophys. Res.* **86**, 8777–8783 (1981).
4. Hunt, G. E., Müller, J.-P. & Gee, P. *Nature* **295**, 491–494 (1982).
5. Jepsen, P. *et al. J. Br. interplanet. Soc.* **33**, 315–320 (1980).
6. Hunt, G. E. *et al. Phil. Trans. R. Soc.* (in the press).
7. Sromovsky, L. A., Revercomb, H. E., Suomi, V. E., Limaye, S. S. & Krauss, R. J. *J. atmos. Sci.* (submitted).
8. Kaiser, M. J., Desch, M. D., Warwick, J. W. & Pearce, J. B. *Science* **209**, 1238–1240 (1980).
9. Ward, N. R., Hunt, G. E. & Saunders, R. W. *Int. J. Remote Sensing* (in the press).
10. Fraedrich, K., Ruprecht, E. & Tunte, U. *J. appl. Met.* **15**, 1312–1316 (1976).
11. Ingersoll, A. P. *et al. J. geophys. Res.* **86**, 8733–8744 (1981).
12. Holton, J. *An Introduction to Dynamical Meteorology* (Academic, New York, 1973).
13. Newburn, R. & Gulkis, S. *Space Sci. Rev.* **3**, 179–271 (1973).

Enhancement of the number of muon catalysed fusions

L. Bracci & G. Fiorentini

Istituto di Fisica dell'Università di Pisa, and INFN, Sez. di Pisa, 56100 Pisa, Italy

We discuss here the effect of an external electric field on the motion of the $(\mu\alpha)$ ions which are produced in the reaction $(d\mu) \rightarrow (\mu\alpha) + n$. We find that appreciable enhancements of the $(\mu\alpha)$ stripping probability can be achieved for strong fields. This effect can be useful for enhancing the number of nuclear fusions which a muon can catalyse in a deuterium–tritium mixture.

The possibility of using the muon catalysis of nuclear reactions for the purpose of energy production has been a long-standing dream of particle physicists¹. Recently, Petrov² presented a scheme which, although based on some optimistic assumptions, can yield a positive energy balance. This scheme involves the muon catalysis of $d + t \rightarrow \alpha + n$ reactions and is supported by the recent discovery³ of a very high formation rate of the $d\text{--}\mu\text{--}t$ molecular ion, $\lambda_{d\mu t} \geq 10^8 \text{ s}^{-1}$ at liquid hydrogen density. Indeed, the theoretical analysis of the process, based on the so-called resonant mesomolecule formation^{4,5}, allows for the possibility of extremely high values of $\lambda_{d\mu t}$.

In the $d\text{--}t$ mixture, the most serious limitation to the efficiency of the fusion chain is the possibility that the muon could stick to the α particle produced in the nuclear reaction, the sticking probability being $W = 1.2\%$ (refs 6, 7). The kinetic energy of the $(\mu\alpha)$ ion is $T_{in} = 3.5 \text{ MeV}$, which corresponds to a velocity $v_{in} = 5.8 e^2/\hbar$. During the slowing down process there is a chance for the $(\mu\alpha)$ to be stripped, so that the muon can take part again to the fusion chain. At a density $\rho_0 = 10^{19} \text{ atom cm}^{-3}$ a fraction $R = 0.20$ of the stucked muons are reactivated, so that the effective sticking probability is $W_{ef} = W(1 - R) = 0.96\%$ while at $\rho = 10\rho_0$ one has $R = 0.21$ and $W_{ef} = 0.95\%$ (see refs 6, 7). The reason for such a limited gain (as depicted in Fig. 1) is that the stripping rate λ_{st} becomes very small as the $(\mu\alpha)$ reaches low velocity, $v \leq e^2/\hbar$. To improve the value

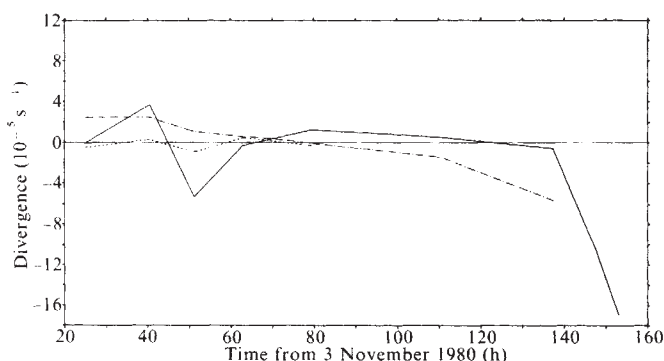


Fig. 3 Measurements of divergence of cloud features in the northern hemisphere of Saturn for 150 h, commencing at 00 h, 3 November 1980. The convective clouds are features 1 (---); feature 2 (—); ····, the UV spot.

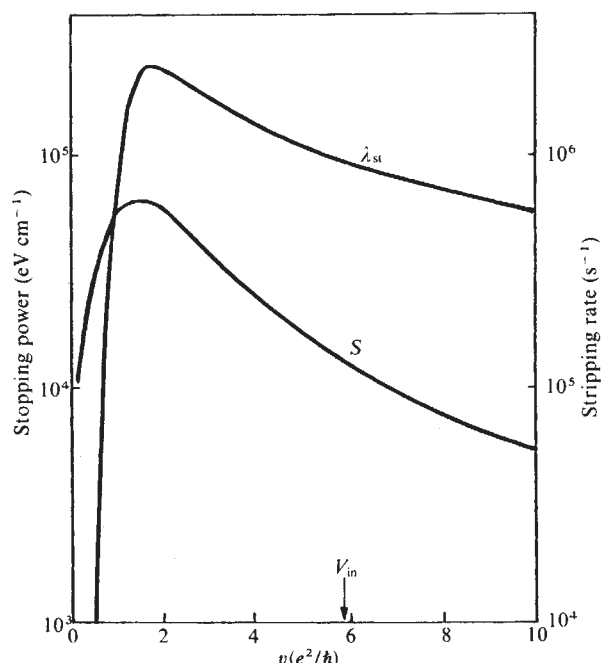


Fig. 1 Stopping power S in eV cm^{-1} and stripping rate λ_{st} in s^{-1} versus $v\hbar/e^2$ at density $\rho = 10^{19} \text{ atoms cm}^{-3}$.

of the reactivation coefficient R the $(\mu\alpha)$ should be kept as long as possible in the velocity range where the stripping rate is high. In principle, this can be achieved by making up for the energy loss due to the stopping power $S(v)$ of the medium by means of an applied electric field.

Clearly the strength E of the electric field which is needed for this purpose is of the order $E \approx S(v_{in})/e$, which means $E \approx 15 \text{ kV cm}^{-1}$ at density $\rho = \rho_0$. This is a rather large value and consequently one has to keep the target density as low as possible. On the other hand, lower limits to the value of the density are required by the condition that the processes relevant to the fusion chain (muon slowing down, atomic capture and atomic cascade, mesomolecule formation) are fast compared with the muon decay rate, $\lambda_\mu = 4.5 \times 10^5 \text{ s}^{-1}$. We assume the mesomolecule formation to be fast at $\rho \geq \rho_0$, which is optimistic but consistent with the present experimental observation and needs comparison with future measurements of $\lambda_{dt\mu}$. In this

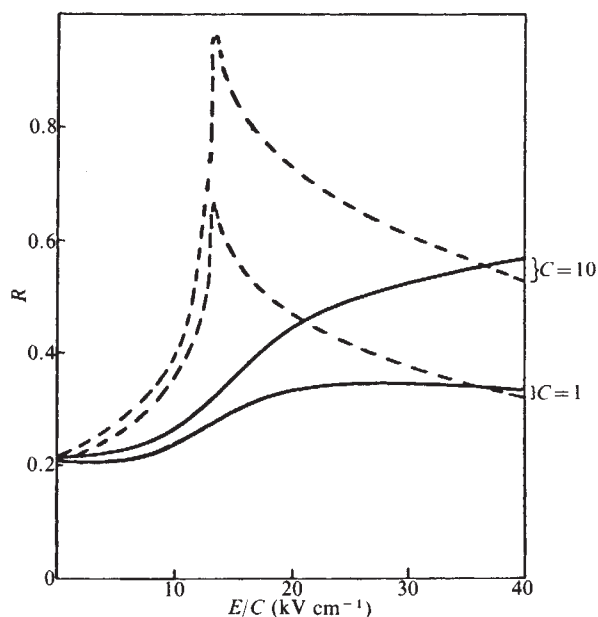


Fig. 2 Reactivation coefficient R versus E/C for the parallel case (dashed line) and the isotropic case (solid line).

way one gets $C = \rho/\rho_0 \geq 1$. We consider two cases, $C = 1$ and $C = 10$.

To calculate the reactivation coefficient we solve numerically the equations of motion of the $(\mu\alpha)$ under the combined action of the electric field and of the dissipative force $S(v)$. This allows us to solve the evolution equation for the number of stripped $(\mu\alpha)$ s. We first consider the case of a static, uniform electric field E : a field that is constant as long as the muon is present in the target and in the region travelled by the $(\mu\alpha)$. One can envisage several configurations depending on the geometry of the apparatus. Of course, the value of R is configuration dependent. We focus on two extreme cases: first where the initial velocity of all the $(\mu\alpha)$ s is aligned with the electric field and second where the initial velocity is randomly oriented. The results, presented in Fig. 2, allow a straightforward interpretation. The peak for the parallel case corresponds to the critical value of the field $E/C = 13.1 \text{ kV cm}^{-1}$, when the dissipative force $S(v_{in})$ is exactly balanced by the electric field. In this case the $(\mu\alpha)$ velocity is constant and the reactivation probability is $R = \lambda_{st}(v_{in})/(\lambda_{st}(v_{in}) + \lambda_\mu)$. In a higher electric field the $(\mu\alpha)$ s are accelerated and therefore the stripping probability is diminished, as is clear from Fig. 1. In lower electric fields the $(\mu\alpha)$ s

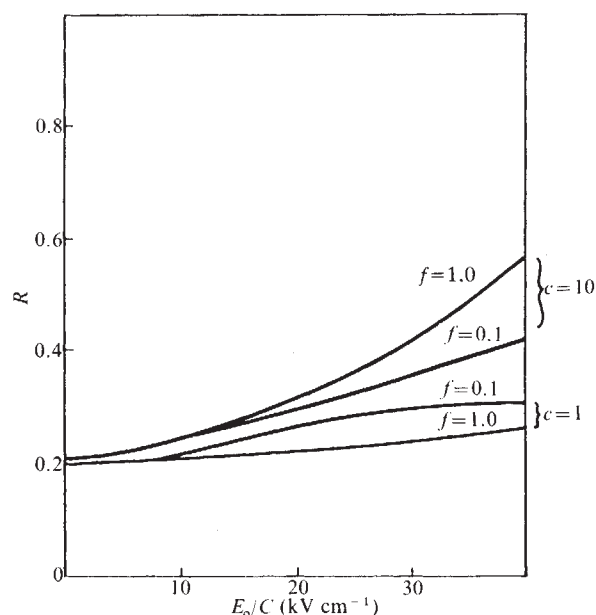


Fig. 3 Reactivation coefficient R versus E_0/C for different values of $f = \omega/\omega_0$ with $\omega_0 = 10^7 \text{ s}^{-1}$.

are slowed down but it takes longer to reach thermal energy than in the zero-field case and consequently $R(E) > R(0)$. Lower values of R are generally found in the isotropic case because for a fraction of the $(\mu\alpha)$ s the slowing down time is shortened by the electric field. In any case there are significant enhancements of R for sufficiently high electric fields.

Another situation of interest is that of an oscillating electric field, $E = E_0 \cos(\omega t + \varphi)$. As the phase (φ) of the field at the time of the $(\mu\alpha)$ formation ($t = 0$) is random, one needs to take a statistical average over φ for the calculation of the reactivation probability. This is equivalent to averaging over the initial orientations of the $(\mu\alpha)$ velocity and consequently one expects the geometry of the apparatus to be irrelevant for the reactivation effect. For this reason we only consider the isotropic case. The qualitative features of the $(\mu\alpha)$ motion are as follows. The velocity is exponentially unstable near the initial value, with a time constant $\tau \approx M/S'(v_{in})$ where M is the $(\mu\alpha)$ mass and S' is the derivative of the stopping power with respect to the velocity. In the range of field strength and frequency of interest (see Fig. 3) this results in a slowing down of the $(\mu\alpha)$. The time the $(\mu\alpha)$ spends with $v \geq e^2/\hbar$ is a decreasing function of ω because, as soon as the electric field becomes opposite to v , the

$(\mu\alpha)$ is quickly decelerated. Next, for $\omega > 2\pi\lambda_\mu$ the $(\mu\alpha)$ velocity can perform several oscillations around zero, with an amplitude $A = eE_0/(\omega^2 M^2 + S'(0)^2)^{1/2}$ before muon decay. In this stage, for a strong electric field and a high density, the $(\mu\alpha)$ spends some time in a region where the stripping rate is appreciable. These considerations allow an easy explanation of the results depicted in Fig. 3. The opposite behaviour with respect to ω at $C = 1$ and $C = 10$ arises from the different contributions of the slowing down stage and of the oscillatory regime to the reactivation probability R .

We conclude that one can enhance the reactivation efficiency by supplying energy to the $(\mu\alpha)$'s by means of an electric field, provided that it is strong enough. On the other hand severe limitations to the strength of the electric field arise because the spark breakdown, is initiated for $E/C \approx 3 \text{ kV cm}^{-1}$ (ref. 8). Thus pulsed and oscillatory field seem preferable, because there it is possible to reach higher peak values than with static fields.

An interesting case would be to use the oscillating electric field in conjunction with a magnetic field B . At the cyclotron frequency $\omega_c = eB/Mc$ one gets the same situation as with a static electric field. Furthermore, the magnetic field can be useful in inhibiting the spark breakdown.

We feel that this approach, despite technical difficulties, can be a help to bring muon catalysed fusion closer to practical applications.

We thank F. Kottman, L. Olivo and, particularly, G. Zweig for helpful discussions.

Received 14 October 1981; accepted 16 March 1982.

1. Alvarez, L. W. *Advent. exp. Phys.* **12** (1972).
2. Petrov, Yu. V. *Nature* **285**, 466 (1980); *Proc. XIV LNPI Winter school*, Leningrad (1979).
3. Bystritsky, V. M. *et al.*, *Phys. Lett.* **94B**, 476 (1980).
4. Vesman, E. A. *Soviet Phys. JEPT Lett.* **5**, 91 (1967).
5. Gershtein, S. S. & Ponomarev, L. I. *Phys. Lett.* **72B**, 80 (1977).
6. Gershtein, S. S. *et al.* *Zh. Eksp. Teor. Fiz.* **80**, 1690 (1981).
7. Bracci, L. & Fiorentini, G. *Nucl. Phys. A* **364**, 383 (1981).
8. Alston, *High-Voltage Technology Atomic Energy Research Establishment*, Harwell.

Search for superheavy elements in monazites using chemical enrichment

R. Stakemann, R. Heimann, G. Herrmann, G. Tittel & N. Trautmann

Institut für Kernchemie, Universität Mainz, D-6500 Mainz, FRG

Evidence for the existence of superheavy elements in monazite inclusions embedded in Madagascan mica and surrounded by giant radioactive haloes was given by Gentry *et al.*¹ who observed photons with energies corresponding to predicted $L_{\alpha 1}$ X-ray energies of element 126 (at 27.25 keV) and also of elements 116, 124 and 127 in irradiations of such crystals with collimated proton beams. For an unambiguous identification, the detection of further members of the L X-ray series would be most important. In X-ray spectra of monazite samples these transitions are buried under the strong K X-ray peaks of the lanthanide elements. They should, however, become visible after chemical enrichment of the superheavy elements². Such a treatment would also solve background problems due to a γ ray at 27.2 keV from the $^{140}\text{Ce}(p,n)^{140}\text{Pr}$ nuclear reaction³⁻⁶ and to K X rays of ordinary trace elements⁷. Although our search remained negative we wish to report the results because our approach differs from that followed by others^{3,8-14} in experiments which also remained negative or did not provide compelling evidence^{6,15}.

Due to a lack of monazite inclusions surrounded by giant haloes we have used bulk monazites as there is no obvious geochemical reason that superheavy elements should not occur in such material if they exist in inclusions although their concentration may be considerably lower. First, we inspected a

Table 1 Concentrations of some trace elements in monazites

Locality	Type of sample	Concentration† (p.p.m.)			
		In	Sn	Sb	Te
Madagascar	Bulk	0.8	10	0.2	0.15
	Crystal*	0.3	1.0	1.2	≤ 0.03
Nigeria	Sand	0.9	350	0.3	0.1
Australia	Sand	0.7	12	1.2	0.9

* Collected³ in the region of Ambatofotsikely where giant haloes occur in the surrounding biotite.

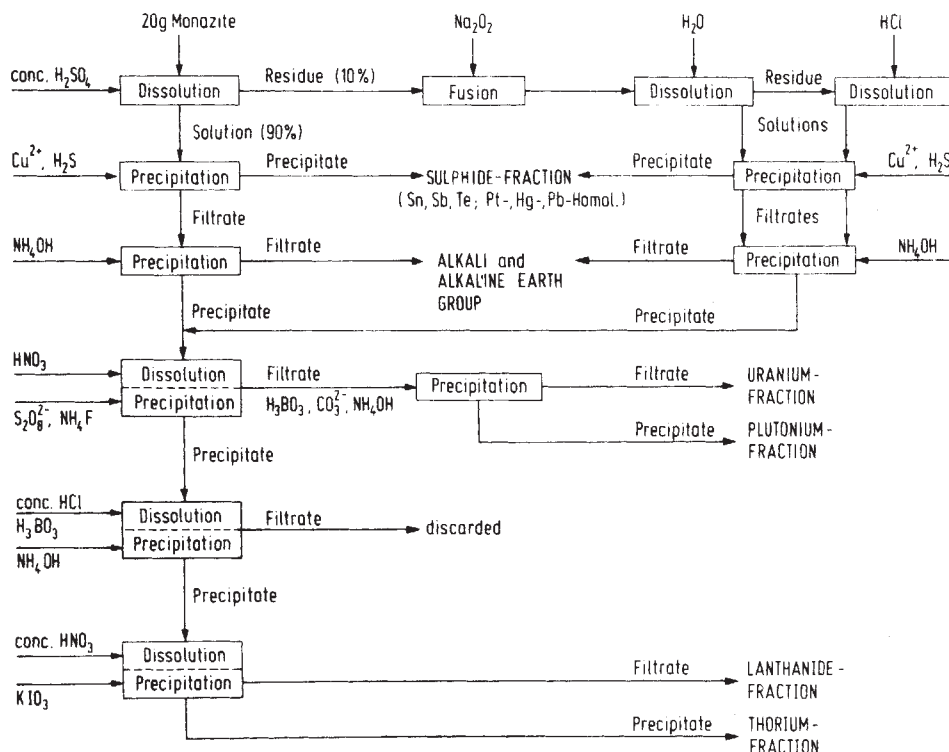
† Average values; due to inhomogenities individual values can vary up to one order of magnitude.

variety of samples, 18 monazites (from Madagascar, Nigeria, South Africa, Malaysia, Singapore, Thailand, Australia, Brasilia, Uruguay and Norway) and 2 bastnasites (from Madagascar), for superheavy element X rays by photon-induced X-ray fluorescence using the 60-keV γ ray of ^{241}Am for excitation. When low-energy γ radiation rather than protons are used for fluorescence excitation, interferences by nuclear reactions can be ruled out. 35-mg samples of the material were mounted on filter paper and exposed to an annular ^{241}Am source which was covered with a 0.05-mm thick rhodium foil to absorb the interfering weak 26.35-keV γ radiation of ^{241}Am . The X-ray spectra were measured with a Si(Li) detector having a resolution of 0.175 keV at 5.9 keV. Two spectra were taken per sample, one for 20 min to determine main constituents, and one for 10 h for trace elements. In the longer exposures, spectra with good statistical quality (about 10^4 counts per channel) were obtained in the most interesting region between 23 and 29 keV.

No evidence for a peak at 27.25 keV energy was found in any of the samples. This corresponds to a concentration limit of ~ 30 p.p.m. for element 126 in bulk monazites and also in bastnasites, a mineral which strongly enriches natural plutonium¹⁶, the most likely chemical homologue of element 126 (ref. 17 and B. Fricke, personal communication). This is about the same limit (6–45 p.p.m.) placed by the proton-induced X-ray fluorescence analysis of monazite inclusions⁶. Clearly present in many samples was the tin $K_{\alpha 1}$ X-ray peak; some samples also showed X-ray peaks of indium and antimony.

To improve the detection limit significantly, six different samples of monazites (from Madagascar, Nigeria, Malaysia and Australia) and one sample of bastnasite (from Madagascar) were selected for chemical treatment. As the chemical behaviour of elements belonging to the superactinide series with 6f–5g electrons cannot be predicted with certainty¹⁷ the procedure applied takes into account that the elements searched for may follow the lanthanides, thorium, uranium or plutonium; even the unlikely case that they form insoluble sulphides or accompany alkaline earth and alkali elements has not been excluded *a priori*. The procedure adopted is shown in Fig. 1. It essentially consists of the following steps: (1) 20-g samples were dissolved in hot concentrated sulphuric acid. The insoluble part was fused with sodium peroxide, dissolved in water and hydrochloric acid and combined with the main sample. (2) Sulphides were co-precipitated with copper sulphide from acid solution; this fraction should also contain superheavy elements around element 114, eka lead. (3) By a hydroxide precipitation the bulk of the cationic elements was separated from alkaline earth and alkali elements. (4) After dissolution of the hydroxides uranium- and plutonium-like elements were oxidized and lanthanides and thorium were precipitated as fluorides. (5) The plutonium fraction was precipitated by ammonia in presence of carbonate ions which kept uranium in solution from where it was finally collected. (6) The fluorides were dissolved in hydrochloric acid plus boric acid and transferred by a hydroxide precipitation into a strong nitric acid solution from which thorium iodate was precipitated; the lanthanides were collected in the filtrate. Bastnasites were processed by a very similar procedure. Samples of the different fractions obtained were

Fig. 1 Flow diagram of the chemical procedure applied in the search for superheavy elements in monazites and bastnasites.



mounted for X-ray fluorescence analysis as described for bulk samples.

As an example for the spectra obtained after chemical enrichment, Fig. 2 shows the spectrum of a plutonium fraction. No peaks can be seen at the positions of the $L_{\alpha 1}$ and $L_{\beta 2}$ X-ray transitions of element 126. From such spectra, upper limits of about 5–20 p.p.b. (at 95% confidence level) are estimated for plutonium-like superheavy elements in the monazites investigated. Peaks of main constituents are also absent; hence decontamination factors of $> 10^7$ have been achieved in the chemical enrichment. Similar concentration limits are estimated for the unlikely case that elements around $Z = 126$ would be collected in the sulphide fraction. For the other fractions, the X-ray fluorescence measurements are less sensitive because only a part of the total fraction could be mounted. We estimate upper concentration limits between 1 and 30 p.p.m. for the different

fractions. These limits hold for both monazites and bastnasites and should be considered as crude estimates.

The concentrations of the elements indium, tin, antimony, and tellurium which have K X-ray peaks in the region of interest were determined in some samples by X-ray fluorescence. Tin and antimony were measured in the sulphide precipitate of the procedure Fig. 1. As part of the tellurium(VI) is found in the hydroxide precipitate, this precipitate was dissolved in concentrated hydrochloric acid, and tellurium(VI) was reduced by boiling and then co-precipitated with copper sulphide. Indium was determined in a separate sample after extraction from sulphuric acid into diethyldithiophosphoric acid and back extraction with strong hydrochloric acid. Chemical yields were measured with radioactive tracers. Residues in the dissolution of monazites were treated separately. The results are given in Table 1. With the exception of tin the concentrations are too low to produce peaks in our spectra of unseparated monazites. However, more than one order of magnitude higher concentrations were found for antimony and tellurium in X-ray spectra excited by protons⁷.

In a selected number of fractions, mainly plutonium fractions, α particle spectra were recorded with semiconductor detectors. No evidence for low-energy emitters reported long ago¹⁸ nor for high-energy α particles¹⁹ postulated as an explanation for the giant haloes¹ was found, in accordance with other studies^{20,21}. Inspection of some fractions by mass spectrometry also did not reveal any anomalies.

We thank H. Ahrens, F. Begemann, W. Brüche, A.-K. Kaffrell, N. Kaffrell, J. V. Kratz, H. Menke, W. Stegmann and M. Weber for help, and the Treibacher Werke, K. Chayawatankur, and D. Schwalm for supplying monazite samples. Financial support by the Studienstiftung des Deutschen Volkes (R.S.), the Bundesministerium für Forschung und Technologie and by the Gesellschaft für Schwerionenforschung Darmstadt is gratefully acknowledged.

Received 16 November 1981; accepted 16 March 1982.

- Gentry, R. V. *et al.* *Phys. Rev. Lett.* **37**, 11 (1976).
- Heimann, R. *et al.* *Ges. für Schwerionenforschung Darmstadt A. Rep.*, 73 (1976).
- Bosch, F. *et al.* *Z. Phys.* **A280**, 39 (1977).
- Fox, J. D. *et al.* *Phys. Rev. Lett.* **37**, 629 (1976).
- Fletcher, N. R. *Phys. Rev. Lett.* **38**, 479 (1977).
- Cahill, T. A. *et al.* *Phys. Rev.* **C17**, 1183 (1978).

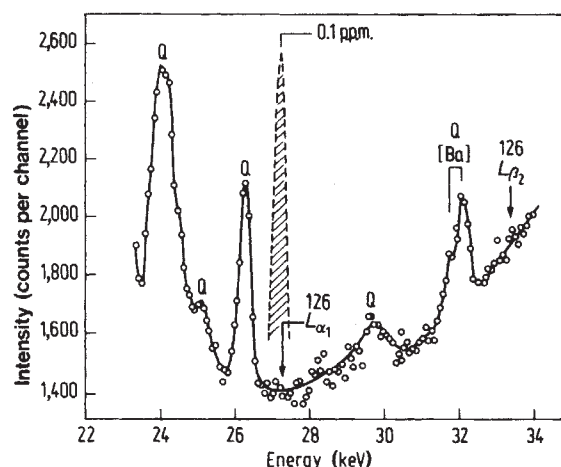


Fig. 2 Photon-induced X-ray fluorescence spectrum of a plutonium fraction isolated from a Madagascan monazite. The hatched peak at the position of the $L_{\alpha 1}$ line of element 126 would result for a concentration of 0.1 p.p.m. in the original bulk monazite sample. Peaks denoted by Q belong to the background stemming from the ^{241}Am photon source and from trace elements (barium) in the filter paper used for sample mounting.

7. Wölfl, W. *et al.* *J. Phys.* **G3**, L33 (1977).
8. Stephan, C. *et al.* *Phys. Rev. Lett.* **37**, 1534 (1976).
9. Ketelle, B. H. *et al.* *Phys. Rev. Lett.* **37**, 1734 (1976).
10. Stoughton, R. W. *et al.* *J. inorg. nucl. Chem.* **41**, 1655 (1979).
11. Sparks, C. J. *et al.* *Phys. Rev. Lett.* **38**, 205, 617 (1977); **40**, 507 (1978).
12. Jelley, N. A. *et al.* *Nature* **265**, 35 (1977).
13. Middleton, R. *et al.* *Phys. Rev. C* **16**, 477 (1977).
14. Annegarn, H. J. *et al.* *Phys. Rev. C* **16**, 379 (1977).
15. Cookson, J. A. *et al.* in *Proc. int. Symp. on Superheavy Elements* (ed. Lodhi, M. A. K.) 164 (Pergamon, New York, 1978).
16. Hoffman, D. C., Lawrence, F. O. Mewherter, J. L. & Rourke, F. M. *Nature* **234**, 132 (1971).
17. Fricke, B. *Structure Bonding* **21**, 89 (1975).
18. Gysae, B. Z. *Naturforschung* **5a**, 530 (1950).
19. Chevallier, A., Chevallier, J., Pape, A. & Debeauvais, M. *J. Phys.* **38**, L331 (1977).
20. Hirdes, D. *et al.* *J. Phys.* **40**, L97 (1979).
21. Fireman, E. L. in *Proc. int. Symp. on Superheavy Elements* (ed. Lodhi, M. A. K.) 172 (Pergamon, New York, 1978).

Environment of Ca^{2+} ions in aqueous solvent

N. A. Hewish*, G. W. Neilson† & J. E. Enderby

H. H. Wills Physics Laboratory, University of Bristol, Royal Fort, Tyndall Avenue, Bristol BS8 1TL, UK

The ion Ca^{2+} is common in nature and has an important role in many biochemical processes¹. However, the way in which this ion behaves in aqueous solution is the subject of much controversy. There are two structural parameters which are necessary for a fuller understanding of Ca^{2+} in solution: the number of water molecules to which Ca^{2+} coordinates in solution, $\bar{n}$, and the configuration which a water molecule adopts relative to it (defined by ϕ , the angle between the plane of the water molecule and the Ca–O axis and r_{CaO} , the Ca–O separation). The technique of neutron diffraction in conjunction with isotopic substitution has been successful in giving unambiguous information regarding $\bar{n}$, ϕ and r for a variety of aqueous solutions². Studies have been carried out on Li^+ , Ni^{2+} and Cl^- at several concentrations³ and Ca^{2+} at one concentration⁴. We report here the results of a study of the structure of the Ca^{2+} – D_2O conformation as a function of concentration.

Investigations have been carried out on several calcium halide and calcium nitrate water systems using X-ray diffraction^{5–9}, and the results have been used to obtain r_{CaO} (ref. 2). However, these experiments have not produced information about the concentration dependence of $\bar{n}$ and ϕ , the central problem which we consider here. The isotopic substitution made was $^{\text{nat}}\text{Ca} \rightarrow ^{44}\text{Ca}$ which changes the neutron scattering length b from 4.90 fm to 1.80 fm.

Neutron diffraction data were gathered at room temperature on the D4 diffractometer of the ILL, Grenoble for two solutions of CaCl_2 in heavy water identical in all respects except for the isotopic state of the calcium (Table 1). Total structure factors, $F(k)$, were derived from the data after correction for multiple scattering and absorption, and were normalized by reference to the scattering from a vanadium standard².

The first-order difference $\Delta(k)$ between the two $F(k)$ s for a given pair of solutions was derived. $\Delta(k)$ is a linear combination of partial structure factors $S_{ij}(k)$ and has the form²:

$$\Delta(k) = A[S_{\text{CaO}}(k) - 1] + B[S_{\text{CaD}}(k) - 1] + C[S_{\text{CaCl}}(k) - 1] + D[S_{\text{CaCa}}(k) - 1]$$

where c_i = atomic fraction of species i ; b_i = coherent scattering length of species i ; $A = 2c_{\text{Ca}}c_{\text{O}} \Delta b_{\text{CaO}}$; $B = 2c_{\text{Ca}}c_{\text{D}} \Delta b_{\text{CaD}}$; $C = 2c_{\text{Ca}}c_{\text{Cl}} \Delta b_{\text{CaCl}}$; $D = c_{\text{Ca}}^2 \Delta b_{\text{Ca}}^2$; $\Delta b_{\text{Ca}} = ^{\text{nat}}b_{\text{Ca}} - ^{44}b_{\text{Ca}}$; $\Delta b_{\text{Ca}}^2 = ^{\text{nat}}b_{\text{Ca}}^2 - ^{44}b_{\text{Ca}}^2$.

Fourier transformation of this function leads to the weighted radial distribution function, $\bar{G}(r)$ given by

$$\bar{G}(r) = A(g_{\text{CaO}}(r) - 1) + B(g_{\text{CaD}}(r) - 1) + C(g_{\text{CaCl}}(r) - 1) + D(g_{\text{CaCa}}(r) - 1)$$

where

$$g_{ij}(r) - 1 = \frac{1}{2\pi^2\rho r} \int (S_{ij}(k) - 1) \sin kr k dk$$

Here $g_{ij}(r)$ is the correlation function relating to the distribution of j -type particles about i -type particles and ρ is the total number density of the solution. $\bar{G}_{\text{Ca}}(r)$ is shown in Fig. 1; the characteristic twin peaks close to 2.4 Å and 3.0 Å are associated with the oxygen and deuterium atoms respectively of the water molecules forming the first coordination shell around the Ca^{2+} ion. Within this first shell, the contribution from the ion-ion

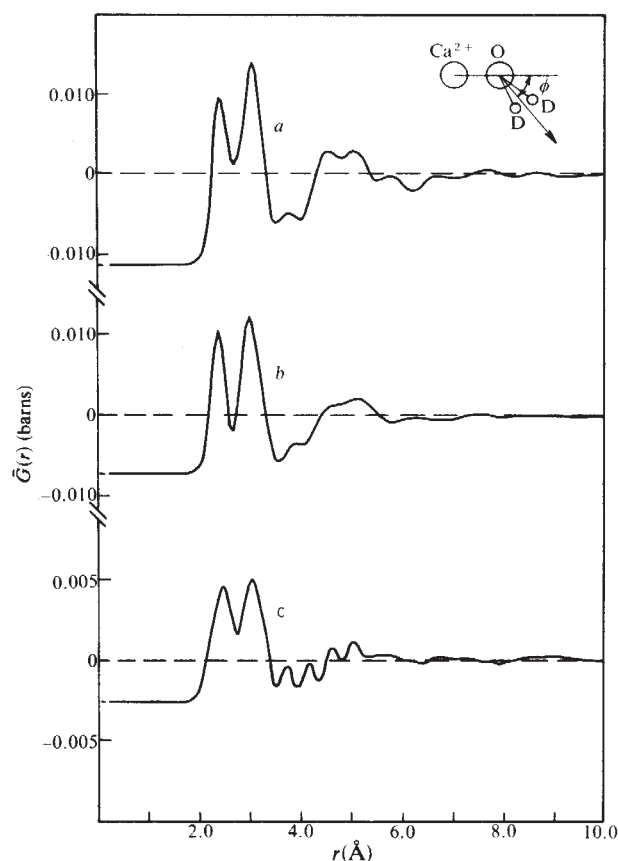


Fig. 1 The weighted distribution $\bar{G}(r)$ for three CaCl_2 solutions in D_2O (a) 4.5 molal (b) 2.8 molal (c) 1 molal. The ion-water geometry is shown at the top right.

distribution functions $g_{\text{CaCl}}(r)$ and $g_{\text{CaCa}}(r)$ are negligible because the coefficients C and D are small relative to A or B (Table 1). The number of atoms, $\bar{n}_i$, to which the areas of any given peak corresponds is calculated using

$$\bar{n}_i = 4\pi c_i \rho \int_{r_1}^{r_2} r^2 g_{\text{Ca}i}(r) dr$$

where r_1 and r_2 are the positions of the minima adjacent to the peak on either side.

The above identification of the twin peaks is confirmed by the calculation of $\bar{n}_{\text{O}}$ and $\bar{n}_{\text{D}}$, in which it is found that $2\bar{n}_{\text{O}} = \bar{n}_{\text{D}}$, within experimental error. The separation of the two peaks may be used to calculate ϕ assuming a geometry in which r_{OD}

* Present address: Standard Telephones Ltd, London Road, Harlow, Essex CM17 9NA, UK.
† To whom correspondence should be addressed.

Table 1 Scattering lengths and sample parameters for CaCl_2 heavy water solution

Molality	c	Isotopes	Scattering lengths (fm)	A	B	C	D
				$(\times 10^{-3} \text{ barns})^*$			
1.0	0.0065	^{40}Ca	4.90	0.77	1.77	0.05	0.01
		^{44}Ca	1.80				
2.8	0.0177	^{40}Ca	4.90	2.03	4.66	0.38	0.07
		^{44}Ca	1.80				
4.5	0.0275	^{40}Ca	4.90	3.02	6.95	0.90	0.16
		^{44}Ca	1.80				

* 1 barn = 10^{-24} cm^2 .**Table 2** Ca^{2+} -water coordination

c (molal)	$r_{\text{CaO}}(\text{\AA})$	$r_{\text{CaD}}(\text{\AA})$	ϕ	$\bar{n}$
1	2.46 ± 0.03	3.07 ± 0.03	$38^\circ \pm 9$	10.0 ± 0.6
2.8	2.39 ± 0.02	3.02 ± 0.03	$34^\circ \pm 9$	7.2 ± 0.2
4.5	2.41 ± 0.03	3.04 ± 0.03	$34^\circ \pm 9$	6.4 ± 0.3

is $1\text{\AA}$, and the angle $\text{D}\ddot{\text{O}}\text{D}$ is 105° . The results of the two latest experiments are summarized in Table 2 and Fig. 1, where we include those derived from an earlier study of a 4.5 molal CaCl_2 .

When we compare these results with those, at comparable concentrations, for a strongly hydrating divalent ion Ni^{2+} (ref. 10) we find the following points of similarity and difference: (1) r_{CaO} , like r_{NiO} does not, when allowance is made for experimental error vary significantly with concentration. In both cases the distances agree within experimental error with those derived from X-ray studies⁵⁻⁹; (2) for both ions at concentrations in excess of 1 molal, ϕ is independent of ionic strength and is $\sim 35^\circ$ which is significantly different from both the 'dipole' configuration ($\phi = 0$) and the 'lone pair' configuration ($\phi = 55^\circ$); (3) for Ni^{2+} , $\bar{n}$ is independent of concentration. For Ca^{2+} , by contrast, $\bar{n}$ is strongly concentration dependent, increasing from ~ 6 to ~ 10 as the molality decreases from 4.5 to 1; (4) the widths of the first peak for both $g_{\text{CaO}}(r)$ and $g_{\text{CaD}}(r)$ decrease with increase in concentration; this behaviour is opposite to that observed for $g_{\text{NiO}}(r)$ and $g_{\text{NiD}}(r)$.

We believe that points (3) and (4) are manifestations of the weakly hydrating nature of Ca^{2+} ions in solution. Unlike Ni^{2+} where our earlier experiments have shown the almost universal nature of the Ni^{2+} - H_2O conformation¹¹ the aqueous environment of Ca^{2+} will depend crucially on the ionic strength, the counter ion and the temperature. It follows that if the properties of Ca^{2+} in solution are to be understood properly a range of experiments of the type we have reported here must be undertaken in which the principal thermodynamic variables (concentration, pressure and temperature) are changed systematically.

N.A.H. thanks the SERC for a research studentship. We all thank the scientific and technical staff at the ILL, Grenoble, and particularly Dr Chieux and Dr Cummings, for valuable help with the experiments. The aqueous solution programme of research is financed by the SERC and we are grateful for their support.

Received 15 February; accepted 16 March 1982.

- Dick, D. A. T. in *Water and Aqueous Solutions* (ed. Horne, R. A.) Ch. 7 (Wiley Interscience, New York, 1972).
- Enderby, J. E. & Neilson, G. W. *Water, a Comprehensive Treatise* Vol. 6 (ed. Franks, F.) 1 (Plenum, New York, 1979).
- Enderby, J. E. & Neilson, G. W. *Rep. Progr. Phys.* **44**, 593 (1981).
- Cummings, S., Enderby, J. E. & Howe, R. A. *J. Phys. C: Solid State Phys.* **13**, 1 (1980).
- van P van Eck, C. L., Mendel, H. & Boog, W. *Disc. Faraday Soc.* **24**, 200 (1957).
- Albright, J. N. *Chem. Phys.* **56**, 783 (1972).
- Licheri, G., Piccaluga, G. & Pinna, G. *J. Chem. Phys.* **64**, 2437 (1976).
- Bol, W., Gerrits, G. J. A. & van P van Eck, C. L. *J. appl. Crystallogr.* **3**, 486 (1970).
- Kollman, P. & Kuntz, I. D. *J. Am. Chem. Soc.* **94**, 9236 (1972).
- Neilson, G. W. & Enderby, J. E. *J. Phys. C* **11**, L625 (1978).
- Newsome, J., Neilson, G. W., Enderby, J. E. & Sandström, M. *Chem. Phys. Lett.* **82**, 399 (1981).

Evolution of passive continental margins and initiation of subduction zones

S. A. P. L. Cloetingh, M. J. R. Wortel & N. J. Vlaar

Vening Meinesz Laboratory, Instituut voor Aardwetenschappen, University of Utrecht, Budapestlaan, 4, 3584 CD Utrecht, The Netherlands

Although the initiation of subduction is a key element in plate tectonic schemes for evolution of lithospheric plates, the underlying mechanisms are not well understood. Plate rupture is an important aspect of the process of creating a new subduction zone, as stresses of the order of kilobars are required to fracture oceanic lithosphere¹. Therefore initiation of subduction could take place preferentially at pre-existing weakness zones or in regions where the lithosphere is prestressed. As such, transform faults^{2,3} and passive margins^{4,5} where the lithosphere is downflexed under the influence of sediment loading have been suggested. From a model study of passive margin evolution we found that ageing of passive margins alone does not make them more suitable sites for initiation of subduction. However, extensive sediment loading on young lithosphere might be an effective mechanism for closure of small ocean basins.

The state of stress at a passive margin is determined by its local geometrical and rheological lithospheric properties and by the system of forces acting on the lithosphere. Of these features the thickness of the oceanic lithosphere⁶, its rheological stratification⁷, the push exerted by the elevation of the oceanic ridge^{8,9} and the forces associated with the (negative) buoyancy of the lithosphere¹⁰ are a function of the age of the oceanic lithosphere. The sediment loading capacity of oceanic lithosphere increases with age, through its continued cooling and densification. One might therefore expect a coupling between the height of the sedimentary column deposited at the passive margin and the age-dependent thermal subsidence of the underlying oceanic lithosphere¹¹. In previous studies of the state of stress at passive margins¹²⁻¹⁴ the possible implications of age-dependent properties of the oceanic lithosphere were not taken into account. We have studied the interrelations between age-dependent forces, geometry and rheology, to decipher their net effect on the state of stress at passive margins. We have constructed finite element models for a passive continental margin in different stages of evolution at ages of 20, 30, 60, 100 and 200 Myr. For all models a half spreading rate of 1 cm yr^{-1} is taken, characteristic of oceanic lithosphere without downgoing slabs attached to it¹⁵. The model features are summarized in Fig. 1. As a model for sedimentary loading we adopt triangular wedges at the continental shelf and rise. As our reference we assume that the maximum thickness of the sediments at the margin corresponds with the thickness that results if the sedimentation has been keeping up with the subsidence of a boundary layer model of the cooling oceanic lithosphere^{11,16}. This implies for the maximum height of the sedimentary triangular wedge an increase from 3.6 km at 20 Myr to 9.4 km at 200 Myr, following roughly a square root of age relation. Observational data on thicknesses of post-rift sedimentary sequences are now available from geophysical surveys carried out on passive margins during the past few years¹⁷. From these data it follows that the reference model is representative of most of the sediment loading histories and resulting thicknesses observed at passive margins. The huge sediment accumulations at deltas, however, clearly exceed the thicknesses as given by the reference model. Therefore, a second class of models was constructed in which the full loading capacity of oceanic lithosphere was taken up by sediments with the sedimentary thicknesses ranging from 10 to 15.8 km. The width of the

sedimentary wedges varies between 150 km for very young margins to 250 km for mature margins, based on data compilations¹⁷. As the sediments replace water we consider only excess densities ($\Delta\rho = 2.4 - 1 = 1.4 \text{ g cm}^{-3}$).

Studies on seismicity and flexure of oceanic lithosphere bending at trenches^{18,19} and analysis of flexure of the oceanic lithosphere under the influence of seamount loading²⁰ have demonstrated that the lithosphere is capable of supporting stresses of the order of some kilobars on geological time scales. The experimental work of Goetze and co-workers^{21,22} shows that the strength of the lithosphere first increases with depth and then decreases rapidly to a level of a few hundred bars in the lower part of the plate, where temperature effects become dominant and the maximum stress achievable is limited by thermally activated ductile flow. Deriving temperature profiles from Crough's²³ model for the oceanic lithosphere, which model combines the merits of the boundary layer model and the plate model, we have constructed lithospheric strength envelopes for a wide range of ages. These are based on Goetze's ductile flow laws in olivine for power-law and Dorn-creep rheologies, assuming a strain-rate of $\dot{\epsilon} = 10^{-18} \text{ s}^{-1}$. For the sediments we assume a zero strength. A similar approach was followed by Bodine and co-workers²⁴. Yield envelopes for lithospheric ages of 30 and 100 Myr are given in Fig. 2 which shows that both the thickness H of the 'mechanical plate' (given by the depth below which the strength of the plate is $<0.5 \text{ kbar}$) and the maximum strength increase strongly with age. The flexure of the lithosphere is counteracted by isostasy. Isostatic forces proportional to the deflection due to loading are included in the model. Of the plate tectonic forces implemented in the models those associated with the ridge push and the negative buoyancy of the oceanic lithosphere are calculated on the basis of Oxburgh and Parmentier's²⁵ model for the formation of oceanic crust and Crough's²³ model for the thermal evolution of oceanic lithosphere. We ignore drag at the base of the lithosphere. Zero horizontal displacements are prescribed for the right-hand boundary of the model to simulate a ridge push transmitted through the continent from an adjacent oceanic plate.

The deformation of the lithosphere at passive margins and the resulting stress field (order of magnitude of a few kilobars) are dominated by sediment loading²⁶. The contribution of the plate tectonic forces to the stress field is an order of magnitude smaller. Differential stresses are largest at the points of maximum flexure. These are located under the rise in the oceanic plate close to the transition of oceanic and rift-stage lithosphere. Figure 2a, b shows the stress maxima for 100 and 30 Myr for the reference load. The lowermost part of the mechanical plate is in yield due to the tensile stresses developed at the base. The main part of the plate remains in the elastic state. The effect of the rheological stratification of the lithosphere is twofold: stress relaxation in the lowermost part and stress concentration in the mechanically stronger upper part. This effect is particularly important if the full loading capacity of the lithosphere is taken up by sediments. This can even result in complete failure of the lithosphere as demonstrated in Fig. 2c.

Figure 2 also shows that the state of stress depends on the age of the margin. To illustrate this dependence more clearly we have plotted in Fig. 3a the maximum differential (tensional)

stresses as a function of age. The age dependence is strongest for ages below 100 Myr. From 30 to 100 Myr—an interval in which the sedimentary loading, the mechanical thickness and the strength increase—the differential stress maxima increase with age. From 100 to 200 Myr the mechanical thickness and the strength of the plate show only a small increase. For ages

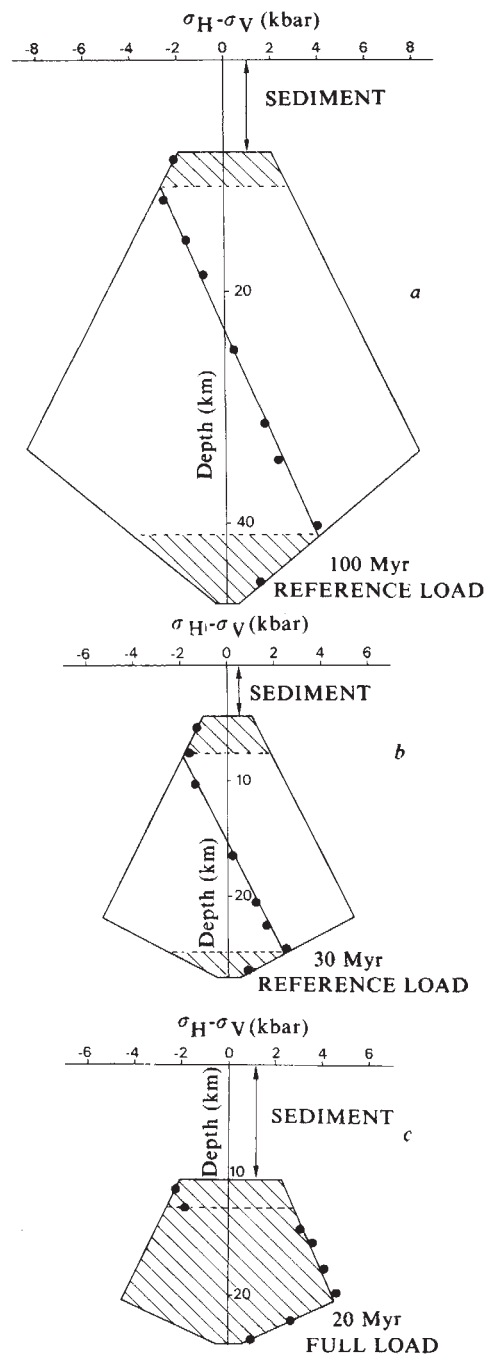


Fig. 2 Strength envelope and results of model calculations (●) at cross-section through the upper part of the lithosphere at the point of maximum flexure (AB of cross-section ABC in Fig. 1). The line inside strength envelope connecting the solid dots is the stress distribution. Differential stresses ($\sigma_H - \sigma_V$) are plotted versus depth. Sign convention for the stresses: tension positive, compression negative. Zero-strength has been assumed for the sediments. Hatched areas in the upper and lower part of the mechanical layer denote failure by brittle fracture and ductile flow respectively. a, b, The results for the reference model of sediment loading for ages of 100 and 30 Myr. c, The results for the full load model for 20 Myr. The horizontal dashed line indicates the neutral surface just before complete failure of 20-Myr old lithosphere takes place.

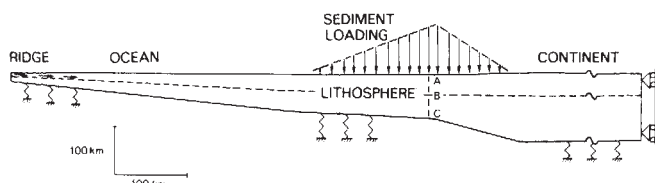


Fig. 1 Model features: geometry, rheology, system of forces and boundary conditions. Mechanical thickness indicated by broken horizontal line. Young's modulus $E = 7 \times 10^{10} \text{ N m}^{-2}$ and Poisson's ratio $\nu = 0.25$.

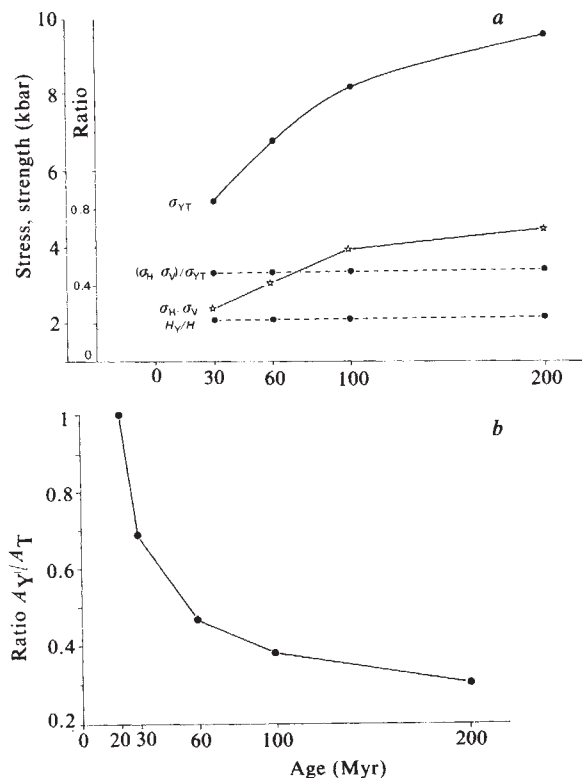


Fig. 3 *a*, Model results for reference load: maximum differential stresses ($\sigma_H - \sigma_V$), tensile strength σ_{YT} , their ratio and the relative thickness of the mechanical layer in failure (H_Y/H) plotted as a function of lithospheric age. *b*, Results full load model: ratio of A_Y (the hatched area inside the strength envelope, see Fig. 2) and A_T (total area of the envelope) as a function of lithospheric age. For ages below 20 Myr complete lithospheric failure is induced.

above 100 Myr the increase in sediment loading according to our reference model results in only a minor increase of the stresses. Interesting quantities are the ratio of the maximum stress generated and the maximum strength, and the ratio of A_Y , corresponding with the hatched area (inside the strength envelope (Fig. 2)) and the total area A_T of the envelope. For the reference model of sediment loading these quantities prove to be essentially independent of age because load, strength and plate thickness all exhibit the same square root of age behaviour due to the thermal evolution of the plate. The situation changes drastically if the full loading capacity is taken up. The surplus load of sediments added to the reference load is most effective in creating high stresses when deposited on a young margin. Figure 3*b* and results of numerical calculations made for ages below 20 Myr (not shown here) demonstrate that full loading on a young passive margin (age below 20 Myr) leads to complete failure of the lithosphere. Owing to the stabilizing affect of the density changes accompanying the formation of oceanic crust²⁵ oceanic lithosphere is gravitationally stable for ages <20–40 Myr (ref. 16). As a result of its further cooling and thermal contraction the lithosphere becomes unstable for ages above 20–40 Myr (refs 16, 25). Taking into account only the lithospheric instability it was reasonable to expect that chances for initiation of subduction of oceanic lithosphere would increase with the age of the margin¹⁰. Our work shows that if after a short evolution of the plate, subduction has not yet started, continued ageing of the passive margin alone does not result in conditions more favourable for plate rupture and initiation of subduction.

Although most of the present deltas of the world are deposited on old oceanic lithosphere²⁷ and very few examples of modern thick sedimentary cones on young lithosphere are known, extensive sedimentary loading might have been an effective mechanism for closure of small ocean basins in geological history. Closure of small oceanic basins has an important

role in the process of mountain building. In this context, note that in a comprehensive account of the evolution of the central Alps, Frisch²⁸ presents evidence for closure of oceanic basins within the first 100 Myr after opening.

For older margins in general, however, it seems that the stresses generated are insufficient to induce lithospheric failure and initiation of subduction. Therefore, pre-existing weakness zones in oceanic lithosphere might be more suitable sites for initiation of subduction than passive margins. This view is consistent with the results of a survey of recently initiated subduction zones in the Pacific²⁹ showing that zones initiated in the Neogene are either at the sites of transform faults or rejuvenated pre-existing subduction zones. Many of the present circum-Pacific zones are in fact the successors of subduction zones already present in the configuration before the breakup of Gondwanaland. Further back in geological time different plate configurations and a different thermal regime might have provided conditions more suitable for initiation of subduction. As an extreme example one might consider the Archaean where, due to the much steeper temperature profiles³⁰, the lithosphere must have been considerably weaker than nowadays. In such a situation plate rupture requires a considerable lower stress level. This suggestion is corroborated by geological data on Precambrian orogenic belts³¹, showing extensive activation of passive margins bordering small oceanic basins.

In the present lithospheric system special circumstances may arise in the process of global plate reorganization, when plate tectonic forces may be concentrated locally. This can be important in an oceanic plate attached to a subduction zone where the pull acting on the subducting slab can be concentrated to a high level (order of magnitude of several kilobars)³². In particular Wortel and Cloetingh³² have shown that lateral variations in the age of the slab descending in a subduction zone might provide a mechanism for fragmentation of oceanic plates and the formation of spreading centres. In an opening oceanic basin, however, the only plate tectonic force to be concentrated is the ridge push, inducing stresses with an order of magnitude of only a few hundred bars. Although the concentration factor might be quite high³³ the level of the concentrated stresses due to ridge push is by no means comparable with the stresses resulting from sediment loading. Therefore, plate reorganization might take place predominantly by the formation of new spreading ridges. In such a process new subduction zones might be subsequently created at the sites of transform faults already present in the plate, when the new spreading direction has a component perpendicular to the direction of the transform fault.

We thank Gerald Wisse (Delft University of Technology) for support with the finite element calculations, and Joel S. Watkins and Rob van der Voo for critically reading an earlier version of the manuscript.

Received 19 January; accepted 8 March 1982.

1. Kirby, S. H. *J. geophys. Res.* **85**, 6353 (1980).
2. Uyeda, S. & Ben-Avraham, Z. *Nature* **240**, 176 (1972).
3. Dewey, J. F. *Am. J. Sci.* **275A**, 260 (1975).
4. Dietz, R. S. *J. Geol.* **71**, 314 (1963).
5. Dewey, J. F. *Earth planet. Sci. Lett.* **6**, 189 (1969).
6. Parsons, B. & Sclater, J. G. *J. geophys. Res.* **82**, 803 (1977).
7. Caldwell, J. G. & Turcotte, D. L. *J. geophys. Res.* **84**, 7572 (1979).
8. Richter, F. & McKenzie, D. *J. Geophys.* **44**, 441 (1978).
9. England, P. & Wortel, R. *Earth planet. Sci. Lett.* **47**, 403 (1980).
10. Vlaar, N. J. & Wortel, M. J. R. *Tectonophysics* **32**, 331 (1976).
11. Turcotte, D. L. & Ahern, J. L. *J. geophys. Res.* **82**, 3762 (1977).
12. Walcott, R. I. *Bull. geol. Soc. Am.* **83**, 1845 (1972).
13. Bott, M. H. P. & Dean, D. S. *Nature phys. Sci.* **235**, 23 (1972).
14. Turcotte, D. L., Ahern, J. L. & Bird, J. M. *Tectonophysics* **42**, 1 (1977).
15. Forsyth, D. W. & Uyeda, S. *Geophys. J. R. astr. Soc.* **43**, 163 (1975).
16. Wortel, M. J. R. Thesis, Utrecht Univ. (1980).
17. Watkins, J. S., Montadert, L. & Dickerson, P. W. *Am. Ass. petrol. Geol. Mem.* **29** (1979).
18. Chapple, W. M. & Forsyth, D. W. *J. geophys. Res.* **84**, 6729 (1979).
19. McAdoo, D. C., Caldwell, J. G. & Turcotte, D. L. *Geophys. J. R. astr. Soc.* **54**, 11 (1978).
20. Watts, A. B., Bodine, J. H. & Steckler, M. S. *J. geophys. Res.* **85**, 6369 (1980).
21. Goetze, C. *Phil. Trans. R. Soc. A* **288**, 99 (1978).
22. Goetze, C. & Evans, B. *Geophys. J. R. astr. Soc.* **59**, 463 (1979).
23. Crough, S. T. *Nature* **256**, 388 (1975).
24. Bodine, J. H., Steckler, M. S. & Watts, A. B. *J. geophys. Res.* **86**, 3695 (1981).
25. Oxburgh, E. R. & Parmentier, E. M. *J. geol. Soc. Lond.* **133**, 343 (1977).
26. Cloetingh, S. A. P. L., Wortel, M. J. R. & Vlaar, N. J. *Am. Ass. petrol. Geol. Mem.* (in the press).

27. Kinsman, D. J. J. in *Petroleum and Global Tectonics* (eds Fischer, A. G. & Judson, S.) 83 (Princeton University Press, 1975).
28. Frisch, W. *Tectonophysics* **60**, 121 (1979).
29. Karig, D. Preprint, Cornell Univ.
30. Sharpe, H. N. & Peltier, W. R. *Geophys. J.R. astr. Soc.* **59**, 171 (1979).
31. Kröner, A. in *Precambrian Plate Tectonics* (ed. Kröner, A.) 56 (Elsevier, Amsterdam, 1981).
32. Wortel, R. & Cloetingh, S. *Geology* **9**, 425 (1981).
33. Kato, T., Shimazaki, K. & Yamashina, K. *Geophys. J.R. astr. Soc.* **60**, 377 (1980).

Geological evidence against the Shyok palaeo-suture, Ladakh Himalaya

Hakim Rai

Wadia Institute of Himalayan Geology, Dehra Dun-248001, India

Evolution of the Ladakh Himalaya is related to the Cretaceous–Tertiary subduction of the Indian oceanic plate below the southern margin of the Eurasian plate along the Indus suture zone^{1–11}. The plate tectonic reconstruction suggests two stages of subduction due to the presence of a pre-Cretaceous suture along the Shyok valley^{12,13}. This hypothesis is based on there being ophiolite belts in the Indus and Shyok valleys. The Indus suture zone is well established but there have been insufficient data on the Shyok suture to confirm the existence of the palaeo-suture. I have studied the Saltoro hills, in the western part of the Shyok valley, and report here the salient geological features. With a few exceptions, the igneous and sedimentary rocks indicate that the Indus and Shyok sutures had the same Cretaceous–Oligocene history. However, the volcanic eruption during the deposition of post-Oligocene molasse in the Saltoro hills of the Shyok valley is the youngest igneous activity in the whole of the Ladakh Himalaya. Hence the geological evidence described here does not accord with the theory of palaeo-suture.

Since the formulation of the plate tectonic theory the occurrence of the Indus ophiolites has been taken as evidence for subduction in that area. It has been suggested that the Himalayan mountain chain came into existence due to the subduction of the Indian plate under the Asian plate during the Cretaceous–Tertiary period^{1–13}. The problem attained a new dimension when an earlier subduction was proposed along the Shyok valley^{12,13} to the north of the Indus suture zone (Fig. 1). Previous work on the Shyok area has been unsystematic^{14–19}. Note that rock masses, other than the igneous, occur as thrust slices and the rock units do not maintain lithological or structural continuity in this region. Recent workers^{16–19} researching in limited and localized areas, have further confused the geology of this area.

The eastern part of the Saltoro hills comprises igneous and sedimentary rocks ranging from Cretaceous to post-Oligocene. Andesites and basalts (the Shyok volcanics) form the oldest igneous rocks erupted before the deposition of flysch. Pyroxenite and hornblende gabbro have intruded into basalt-andesite association along the northern part of the Saltoro hills and in the lower Shyok valley. The assemblage of basic rocks along with metasedimentary rocks of pre-Cretaceous age—the southern limit of the Karakoram sediments¹⁵—has again been intruded by granitoids, causing thermal metamorphism and migmatization.

Major igneous activity ended in the region with the emplacement of Tertiary batholithic granitoids of the Ladakh and Karakoram²⁰. Sandwiched between these igneous rocks are two distinct sedimentary belts of flysch and molasse. Both these belts have thrust contacts with each other as well as with the igneous rocks and occur as detached masses. The flysch forms a marginal part of the eastern Saltoro hills and separates the basic volcanics from the molasse. It is dominated by shale and limestone with occasional sandstone or quartzite. The rocks are tightly folded near the thrust contact and are generally dipping due north. The limestone, particularly in the lower part of the flysch, is highly fossiliferous and has yielded the following fossils

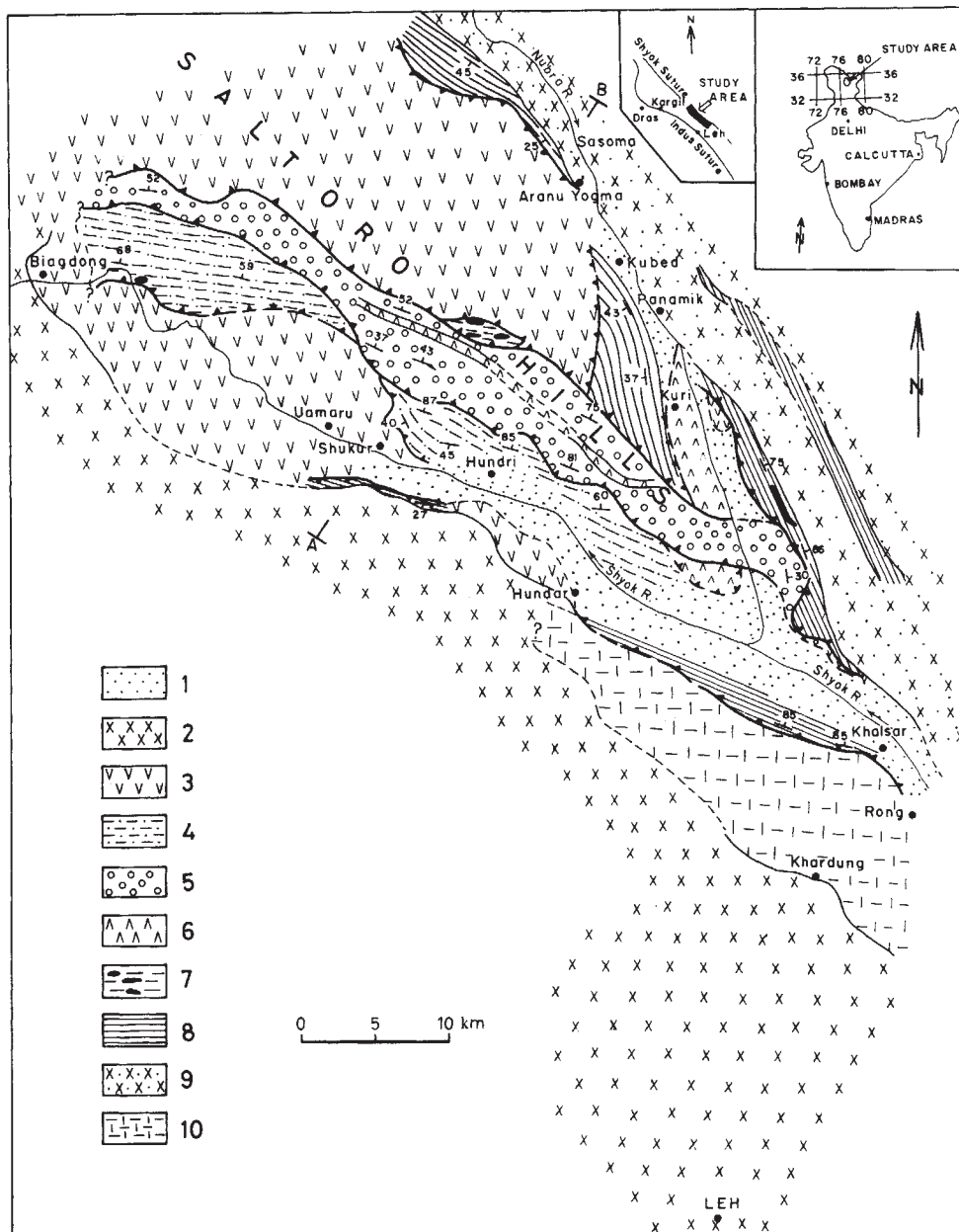
of Upper Cretaceous–Eocene age: *Fasciolites oblonga* (d'Orbigny), *Nummulites* sp. aff. *N. obtusus* Sowerby, *N. ataciscus* Leymerie, *Assilina* sp., *Dictyoconus* sp., *Lockhartia* sp., *Cyclammina* sp., *Cerriocava nilkanthi* Singh, *Feddenia* garkhalensis Mathur, *Lophosmia* sp., *Girvanella* sp., *Lithophyllum* sp. and charophyte gyrogonite. Beside these characteristic fossils, the rocks have yielded turrillid gastropods, brachiopod shells and crinoid stems.

A thin belt of molassic sediments, indicative of shallow continental conditions, contain rounded to sub-rounded pebbles and boulders of limestone, chert, basalt, andesite, rhyolite and granitoids. The limestone boulders have yielded *Orbitolina* sp., *Dictyoconus* sp., *Nummulites* sp., ill-preserved gastropods (turrillides) and so on. The granitic and rhyolitic clasts have been derived from the Tertiary batholiths (Ladakh and Karakoram) and Khardung acid volcanics (38 ± 2 Myr)²¹ of the Shyok valley. Hence the molasse is post-Oligocene. Penecontemporaneous andesitic lava flows and subsequent intrusion of basic and acidic dykes are significant features of this molasse, which is folded and has thrust contacts with flysch in the south and basic volcanics (with associated granitoids) to the north. Emplacement of discontinuous bodies of chromite-bearing serpentinites with black shales has taken place along the thrust contacts, especially along the northern border of the molasse.

The Himalayan mountain chain has been considered to be a typical example of continent–continent collision³. It is thought that the subduction of the north-moving Indian plate started along the Indus suture during the end of the Cretaceous and the actual collision of the continental parts of the two plates probably began in the late Eocene or early Oligocene^{10,22}. Recent studies by Frank *et al.*¹³ in the Shyok valley, to the north of the Indus suture zone, have also included all the volcanics (undifferentiated) with Shyok ophiolites and assigned an age older than the Indus ophiolites. This led them to reconstruct a new plate tectonic model for the Himalayan mountain belt, proposing two stages of subduction separated in space and time. Accordingly, the first subduction was initiated in the pre-Cretaceous time (exact time of subduction is not given) along the Shyok suture, which became inactive later. The second subduction followed the Indus suture during the Upper Cretaceous–Tertiary period. The later predominance of the Khardung acid volcanics younger than the Ladakh Batholith described from the Shyok valley^{17,18} was opposed by Bhandari *et al.*¹⁶ and Thakur *et al.*¹⁹, who established the intrusion of batholithic units into these acid volcanics. Hence, the Khardung volcanics are older than the Ladakh batholith. They also separated the acid and basic volcanic rocks of this region as Khardung volcanics and the Shyok volcanics, respectively.

I have found three distinct volcanic phases in the Shyok valley which differ in composition, age and the environment of eruption. The oldest volcanic rocks, representing the Shyok ophiolites, are basic in composition and have suffered thermal metamorphism near the contacts with gabbro, pyroxenite, tonalite and granite. This complex volcanic–plutonic association is partially overlain by the Upper Cretaceous–Eocene flysch and post-Oligocene molasse, a case similar to the Indus ophiolites. Thus the Shyok valley marks the northern extremity of the Indus ophiolites as there is no evidence to indicate the separation of these two belts during the Cretaceous–Eocene. Extensive occurrence of basic xenoliths within the Ladakh batholith suggests that the area between the Indus and Shyok valleys was occupied by the same ophiolitic rocks which are now associated with the Indus and Shyok ophiolites. These rocks provided a thick cover for the batholithic intrusion. Subsequent erosion has removed the cover rocks over the batholith. Consequently, the Shyok ophiolites apparently form a separate belt that runs parallel to the Indus ophiolite belt. A second volcanic eruption is documented by the Khardung volcanics having acidic composition and constituting a distinct belt along the northern margin of the Ladakh batholith between Hundar and Khardung villages (Fig. 1), and extends eastwards beyond the area under discussion. The batholithic units have intruded into these volcanics

Fig. 1 Geological map of the Shyok valley: 1, river deposits; 2, granitoids of the Ladakh batholith; 3, Shyok volcanics with pyroxenite; 4, Cretaceous-Tertiary flysch; 5, molasse of Saltoro hills; 6, synsedimentary volcanics associated with molasse; 7, detached serpentinite bodies along with black shales; 8, Karakoram sediments (Permian?); 9, granitoids of the Karakoram batholith; 10, Khardung volcanics (acid volcanics).



at several places. The association of lapilli tuffs and conglomeratic beds in the Khardung volcanics suggests that the rocks were erupted in shallow water to subaerial conditions. Lapilli tuffs at the top of the main phase of this eruption are conformably overlain by lahars or mud flows having intermittent lava flows of rhyolitic composition. The lahars are conformably overlain by coarse clastic, non-marine deposits, essentially composed of a sandstone-shale sequence with conglomerates.

The third and the last phase of volcanic activity is preserved within the molasse of the Saltoro hills. The reddish-brown volcanic rock is characterized by random distribution of euhedral plagioclase phenocrysts (An_{36-32}) in the fine groundmass. Dykes of similar composition have intruded into the basic volcanics along the southern flank of Saltoro hills. The molasse with which these andesitic volcanics are associated was deposited in a trough formed within basic volcanics which were intruded by the granitoids of the Ladakh batholith from the south and the Karakoram granitoids from the north.

Because the molasse is rich in granitic material, probably derived from the Ladakh batholith (27.6 ± 0.6 Myr)²¹, pebbles and boulders of acid volcanics (Khardung volcanics = 38 ± 2 Myr)²¹ and limestone containing Palaeocene foraminifera (as in the flysch), it can be correlated with the Kargil molasse of Miocene to Pleistocene age exposed along the Indus suture

zone. In the absence of fossils no exact age can be assigned to the matrix of the molasse. The extrusive and dyke rocks associated with molasse of the Saltoro hills are either Miocene or younger. No such igneous activity has been observed in the molasse along the Indus suture zone. This evidence suggests a prolonged igneous activity in the Shyok valley. The volcanics and dykes of the Saltoro hills appear to be the youngest igneous rocks of the Ladakh Himalaya.

It has been proposed that after subduction along the Shyok suture, a new phase of subduction commenced along the Indus suture zone during the Upper Cretaceous and continued up to the late Oligocene or early Miocene. If the subduction along the Shyok suture stopped before the initiation of the new site of subduction along the Indus suture then the volcanic rocks in the Shyok valley should be much older than Upper Cretaceous. It is

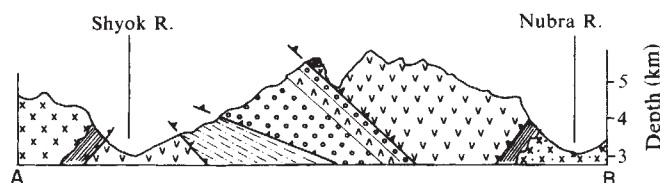


Fig. 2 The geological cross-section along A-B in Fig. 1.

evident from the above that the palaeo-suture idea^{12,13} as proposed does not find any support lithologically, structurally or stratigraphically.

A comparison of the present observations from the Shyok valley with the existing geological data from the Indus suture makes it possible to correlate the geological events of these two regions. There has been widespread volcanic-plutonic igneous activity in the Ladakh Himalaya, extending from the Indus suture zone to the Karakoram during the initiation of the Himalayan orogeny. The emplacement of basic and ultrabasic rocks took place from the Dras to Shyok valley on a regional scale. The Ladakh batholith has intruded into this basic and ultrabasic rock association. The existence of two parallel ophiolite belts is simply a result of batholithic intrusion and subsequent deep erosion, removing the cover rocks of the batholith which is rich in basic xenoliths. Hence the Shyok ophiolites represent the northern extension of the Indus ophiolites emplaced during the Upper Cretaceous–Oligocene.

I conclude that shallow marine conditions existed in the Shyok region at least up to the early Eocene. Although the Indus suture zone became magmatically inactive during the end of the Oligocene, the Shyok valley represents still younger igneous activity. Therefore, the present geological evidence does not favour the two-stage subduction with a palaeo-suture in the Shyok valley.

I thank Professor I. C. Pande, N. S. Viridi and M. I. Bhat for comments, N. S. Mathur for identifying the fossil assemblage and the Director, Wadia Institute of Himalayan Geology, Dehra Dun for permission to publish this paper.

Received 17 December 1981; accepted 16 March 1982.

- Dewey, J. F. & Bird, J. M. *J. geophys. Res.* **75**, 2625–2647 (1970).
- Powell, C. McA. & Conaghan, P. J. *Earth planet. Sci. Lett.* **20**, 1–12 (1973).
- Dewey, J. F. & Burke, K. C. A. *J. Geol.* **81**, 683–692 (1973).
- Crawford, A. R. *Geol. Mag.* **111**, 369–480 (1974).
- Le Forte, P. *Am. J. Sci.* **275A**, 1–44 (1975).
- Menke, H. W. & Jacob, K. H. *Bull. seis. Soc. Am.* **66**, 1695–1711 (1976).
- Bird, P. *J. geophys. Res.* **83**, 4975–4987 (1978).
- Armbruster, J., Seeber, L. & Jacob, K. H. *J. geophys. Res.* **83**, 269–282 (1978).
- Klootwijk et al. *Earth planet. Sci. Lett.* **44**, 47–64 (1979).
- Brookfield, M. E. & Reynolds, P. H. *Earth planet. Sci. Lett.* **55**, 157–162 (1981).
- Mitchell, A. H. G. *J. geol. Soc. Lond.* **138**, 109–122 (1981).
- Gansser, A. *Coll. Inter C.N.R.S. Ecol. Geol. de l'Himalaya* **268**, 181–191 (1977).
- Frank, W., Gansser, A. & Trommsdorff, V. *Schweiz. miner. petrogr. Mitt.* **57**, 89–113 (1977).
- Stoliczka, F. *Mem. geol. Surv. Ind.* **5**, 12–15 (1865).
- Norin, E. *Rep. Sino-Swedish Expedn Publ.* **293**, 340 (1946).
- Bhandari et al. *Abstr. 9th Sem. Himalayan Geology* (1978).
- Sharma, K. K. & Kumar, S. *Him. Geol.* **8**, 262–287 (1978).
- Sharma, K. K. & Gupta, K. R. *Rec. Res. Geol.* **7**, 133–143 (1978).
- Thakur, V. C. et al. *J. geol. Soc. Ind.* **22**, 46–50 (1981).
- Desio, A., Tongiorgi, E. & Ferrara, G. *22nd int. Geol. Cong.* **11**, 479–497 (1964).
- Molnar, P. & Tapponnier, P. *Science* **189**, 419 (1975).
- Sharma, K. K. et al. *Him. Geol.* **8**, 268–295 (1978).
- Rai, H. *Abstr. 10th Sem. Himalayan Geology* (1978).

Tightly bound β -hydroxy acids in a Recent sediment

Kimitaka Kawamura* & Ryoshi Ishiwatari

Department of Chemistry, Faculty of Science, Tokyo Metropolitan University, Fukuzawa 2-1-1, Setagaya-ku, Tokyo 158, Japan

The identification of β -hydroxy acids in Recent sediments^{1–6} is considered to be evidence of a bacterial contribution to sedimentary organic matter. In previous studies, β -hydroxy acids have been extracted by saponification with KOH/methanol under reflux. We have now found that β -hydroxy acids are also present in Recent sediment in a tightly bound form which cannot be extracted by ordinary saponification. Release of those β -hydroxy acids requires that the sediment be heated to ~150 °C before saponification, or the use of harsher saponification.

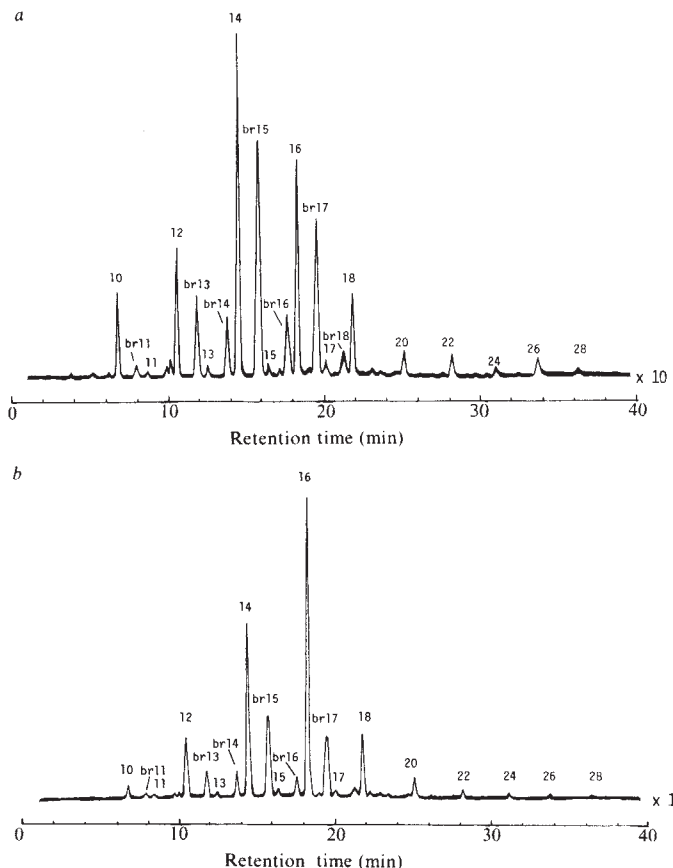


Fig. 1 Mass fragmentogram (m/z 175) of β -hydroxy acid methyl ester TMS ethers from unheated (a) and heated (154 °C; b) sediment samples.

The use of mild heating to study the organic matter content of Recent sediments has been shown to increase the amount of extractable fatty acids^{7–9}. The increase is due to so-called 'tightly bound' fatty acids⁹, which are thought to be located in organic and/or inorganic matrices where saponification reagents cannot easily act, or to be tightly linked in some unknown way with the organic and/or inorganic matrices⁹. These studies suggested that organic molecules other than fatty acids might also be present in sediments in a 'tightly bound form'. As β -hydroxy acids are considered markers of bacterial activity^{1–4}, it would be of interest to identify a tightly bound form of these compounds in sediment.

Surface sediment samples were taken from Lake Biwa, the largest lake (altitude 85 m, area 674.4 km², maximum depth 102 m) in Japan, which is oligotrophic. A wet sediment sample was homogenized using a Nihon-seiki (AM-1) homogenizer and aliquots of the sample (5–10 g wet wt) were heated at various temperatures (68–325 °C) in a sealed Pyrex tube (12 mm i.d. $\times$ 15 cm) under nitrogen for 24 h. The unheated and heated samples were saponified with 50 ml of 0.5 M KOH/methanol containing 5% water under reflux for 2 h. Saponification was carried out only once because we found that repeating the procedure increases the amount of β -hydroxy acids obtained by <3%. The alkaline solution was extracted with 20 ml of *n*-hexane/ether (9:1) to remove neutral components. The remaining solution was then acidified with concentrated HCl and the acid components extracted with *n*-hexane/ether (9:1). The extracts were methylated with 14% BF₃/methanol and the methyl esters separated into monocarboxylic acid fraction with *n*-hexane/benzene (9:1) and hydroxy acid fraction with benzene/ethyl acetate (1:1) by silica gel chromatography (Mallinckrodt 100 mesh; 5 mm i.d. $\times$ 10 cm). The latter fraction was treated with TMS-BA [N,O-Bis(trimethylsilyl)acetamide 25% in acetonitrile: Tokyo Kasei Co. Ltd] reagent to derive TMS ethers of hydroxy acid methyl

* Present address: Institute of Geophysics and Planetary Physics, University of California at Los Angeles, Los Angeles, California 90024, USA.

Table 1 Concentrations of β -hydroxy acids released from the sediment, pre-extracted sediment and isolated kerogen

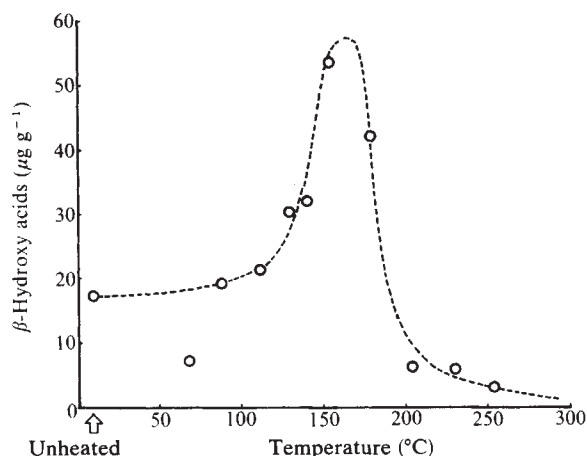
	Sediment*	Pre-extracted sediment	Kerogen
$\mu\text{g per g dry weight sediment}$	36.3	28.7	6.8

* The amount was obtained by subtracting the concentration ($17.2 \mu\text{g per g}$) for unheated sample from the concentration ($53.5 \mu\text{g per g}$) for heated sample at 154°C .

esters, which were then analysed by gas chromatography-mass spectrometry (Shimadzu-LKB 9000) on a glass column (3 mm i.d. $\times$ 2 m) packed with 1% OV-1, with programmed temperature increases of 5°C per min from 100 to 280°C . The β -hydroxy acids were quantified by comparing the peak heights in the mass fragmentograms scanned at m/z 175, which is a characteristic fragment of β -trimethylsilyloxy methyl esters^{1,2}, with authentic standards (β -hydroxy normal C_{14} , C_{16} and C_{18} acids).

β -Hydroxy C_{10} – C_{28} acids were detected in the unheated and heated samples. Mass fragmentograms scanned at m/z 175 for the unheated and heated samples (Fig. 1) showed that most even carbon-numbered β -hydroxy acids are unbranched (normal), whereas odd carbon-numbered acids are almost entirely iso/ante-iso branched chain compounds. A similar distribution has been reported³ for bound β -hydroxy acids from algal mats sediments. As predicted, the amount of both normal and branched β -hydroxy acids increased on heating; in particular, the concentration of normal β -hydroxy C_{16} acid increased by as much as fivefold. With increasing temperature, the total amount of β -hydroxy acids increases threefold and maximizes at 150 – 180°C (Fig. 2). As sedimentary β -hydroxy acids are considered to be produced by bacteria^{1–4}, our observations suggest the presence in the sediment of bacteria which produce specifically a large amount of β -hydroxy n - C_{16} acid. Apart from n - C_{16} acid, the relative proportions of constituents in the unheated and heated samples are similar, suggesting that other acids have a similar origin.

β -Hydroxy acids could not have been formed directly through β -oxidation of saturated fatty acids because we confirmed that authentic n - C_{17} saturated fatty acid did not generate any β -hydroxy acids when heated with wet Ca-montmorillonite at 160°C for 24 h, and because we could not detect any decrease in the concentration of saturated fatty acids in sediment heated to 150 – 180°C . α , β -Unsaturated fatty acids are other possible precursors of β -hydroxy acids. We consider this route unlikely, however, because α , β -unsaturated fatty acids have never been found in sediments^{10,11}. It therefore seems that the β -hydroxy acids released on heating may not be artefacts but may have existed originally in the sediments.

**Fig. 2** Changes in the concentrations of β -hydroxy acids in sediments on heating.

Next, we conducted a harsher saponification (180°C , 3 h) of the pre-extracted sediment (0.5 g) and the isolated kerogen (34.5 mg, ash content 32.5%) using a Pyrex tube (12 mm i.d. $\times$ 15 cm) containing 5 ml of 2 M KOH solution¹². As shown in Table 1, we obtained a considerable amount of β -hydroxy acids. The amount of β -hydroxy acids released from the pre-extracted sediment by harsher saponification ($28.7 \mu\text{g g}^{-1}$) is almost equal to that released from the sediment by heating ($36.2 \mu\text{g g}^{-1}$). β -Hydroxy acids were also released from kerogen, although at a lower concentration ($6.8 \mu\text{g g}^{-1}$ of dry sediment) than from the pre-extracted sediment. Harsher saponification with 2 M KOH solution may dissolve and/or decompose some humic materials, amorphous silica and clay minerals. It is therefore likely that most β -hydroxy acids were associated with those materials: that is, 'tightly bound' β -hydroxy acids can be considered to be occluded by (and partly linked with) humic materials, including kerogen and amorphous silica, and to be trapped in the interlayers of clay minerals. Further work is necessary to define the type of binding.

β -Hydroxy acids have been found in bacteria^{2,5,13–16}, but not in algae. We did not detect these compounds in phytoplankton (*Closterium*) taken from Lake Biwa. Therefore, it is obvious that sedimentary β -hydroxy acids are a result of bacterial activity. As most (68%) of the β -hydroxy acids in the sediment exist in a 'tightly bound form', the quantities of β -hydroxy acids reported previously may be underestimates.

This study was supported in part by a grant from the Ministry of Education, Japan.

Received 25 August 1981; accepted 8 March 1982.

- Eglinton, G., Hunneman, D. H. & Douraghi-Zadeh, K. *Tetrahedron* **24**, 5929–5941 (1968).
- Boon, J. J., DeLange, F., Schuyl, P. J. W., DeLeeuw, J. W. & Schenk, P. A. in *Advances in Organic Geochemistry 1975* (eds Campos, R. & Goni, J.) 255–272 (Enadimsa, Madrid, 1977).
- Cardoso, J. N., Eglinton, G. & Holloway, P. J. in *Advances in Organic Geochemistry 1975* (eds Campos, R. & Goni, J.) 273–287 (Enadimsa, Madrid, 1977).
- Perry, G. J., Volkman, J. K., Johns, R. B. & Bavor, H. J. Jr *Geochim. cosmochim. Acta* **43**, 1715–1725 (1979).
- Volkman, J. K., Johns, R. B., Gillan, F. T., Perry, G. J. & Babor, H. J. Jr *Geochim. cosmochim. Acta* **44**, 1133–1143 (1980).
- Cranwell, P. A. *Geochim. cosmochim. Acta* **45**, 546–552 (1981).
- Baedecker, M. J., Ikan, R., Ishiwatari, R. & Kaplan, I. R. in *Chemistry of Marine Sediments* (ed. Yen, T. F.) 55–72 (Ann Arbor Science, Michigan, 1977).
- Harrison, W. E. *Chem. Geol.* **21**, 314–334 (1978).
- Kawamura, K. & Ishiwatari, R. *Geochim. J.* **15**, 1–8 (1981).
- Van Vleet, E. S. & Quinn, J. G. *Nature* **262**, 126–128 (1976).
- Matsuda, H. & Koyama, T. *Geochim. cosmochim. Acta* **41**, 341–345 (1977).
- Schnitzer, M. & Neyrould, A. *Fuel* **54**, 17–19 (1975).
- Moss, C. W., Samuels, S. B., Liddle, J. & McKinney, R. M. *J. Bact.* **114**, 1018–1024 (1973).
- Mayberry, W. R., Smith, P. F., Langworthy, T. A. & Plackett, P. *J. Bact.* **116**, 1091–1095 (1973).
- Boon, J. J., DeLeeuw, J. W., Hoek, G. J. & Vosjan, J. H. *J. Bact.* **129**, 1183–1191 (1977).
- Mayberry, W. R. *J. Bact.* **143**, 582–587 (1980).

Surface coatings on ancient coccoliths

Peter M. Bürki*, Lesley S. Dent Glasser
& Douglas N. Smith

Department of Chemistry, University of Aberdeen, Meston Walk,
Old Aberdeen AB9 2UE, UK

Coccoliths are rings of calcite plates, often of intricate design, produced by unicellular algae¹ (Fig. 1). Still widespread in modern oceans, these algae must have been extremely abundant in the Cretaceous period, and their skeletal remains form the main component of natural chalk. Both modern and ancient coccoliths are well preserved even in the corrosive environments of seawater and sediments, and the surface properties of natural chalk are said to be different from those of synthetic calcite. It has been suggested that this may be the result of a surface membrane; the present study establishes the existence of a coating on ancient coccoliths, and discusses its nature.

* Present address: Holderbank Management & Consulting Ltd, Materials Division, CH-5113 Holderbank, Switzerland.

It has been demonstrated^{2,3} that modern coccoliths are surrounded by a thin film of organic material. This could be composed of polysaccharides and amino acids⁴ and may be formed when the coccolith is produced within the algal cell⁵. By careful dissolution of Oligocene coccoliths on electron microscope grids, Hamano and Honjo⁶ observed residual membranes which they believed to be composed partly of organic matter. They found that the acid-insoluble residue from the coccolith fraction of the chalk contained polysaccharides, amino acids and clay-like material. Taylor (unpublished work) suggested that thin layers of precipitated clay minerals might cover and protect coccoliths in chalk. This postulate is substantiated by the work of Weir and Catt⁷ and Jeans⁸ who concluded that smectite is precipitated from porewater in chalk sediments. Suess⁹ has also postulated organo-clay associations for all pre-Recent limestones.

In the present study, samples of Cretaceous chalk from Swanscombe were gently decalcified on electron microscope

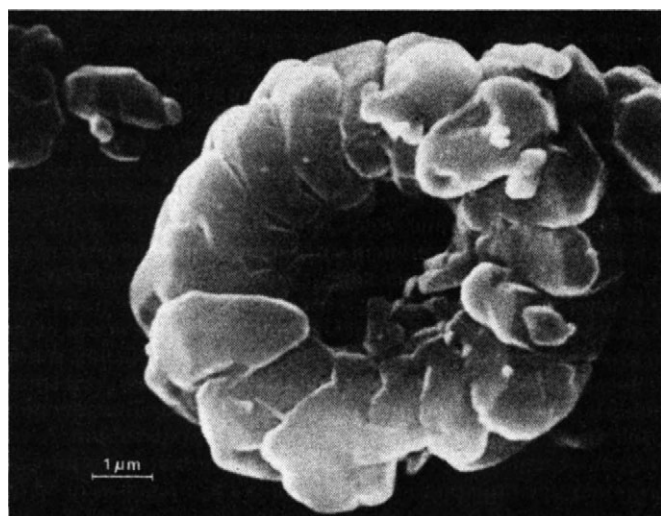


Fig. 1 Scanning electron micrograph of a Cretaceous coccolith. This is only one of the many forms observed.

grids and examined in a CORA analytical transmission electron microscope (Kratos Ltd, AEI Scientific Instruments). Residues of decalcified coccoliths are shown in Figs 2 and 3. These complete 'skins' are clearly replicas of the original structures, an example of which is shown in Fig. 1. Slower dissolution in EDTA buffered with $\text{NH}_4\text{Cl}/\text{NH}_4\text{OH}$ produced both partly and completely decalcified rings, confirming that the coatings are on the surface of the coccoliths. Whole skins, however, were rather rare; distorted fragments either partly retaining the original structure or lacking any recognizable form were more common. The coatings may collapse after dissolution or be ruptured by the production of CO_2 .

Electron microanalysis was difficult because the coatings are very thin (a few hundred ångströms at most) giving a low count rate. By examining areas over holes in the substrate, the problem of background count was avoided, and the approximate atomic ratios found were: 1.0 Si:0.1 Mg:0.3 Al:0.04 K:0.1 Fe. P, from apatite crystals, S and Cl were also occasionally encountered. Significant levels of Ca were present only when visible fragments of calcite remained. These ratios, which were the same with all three dissolution procedures (see legend to Fig. 2), are typical of a smectite, but none of the skins gave a diffraction pattern. Similar ratios were found for the other amorphous debris of ill-defined form. Crystalline clay minerals, quartz and apatite were also observed in the residues, occasionally attached to the coatings (Fig. 3).

Because the microanalyser cannot detect carbon, no conclusion can be drawn about the organic nature of these skins.

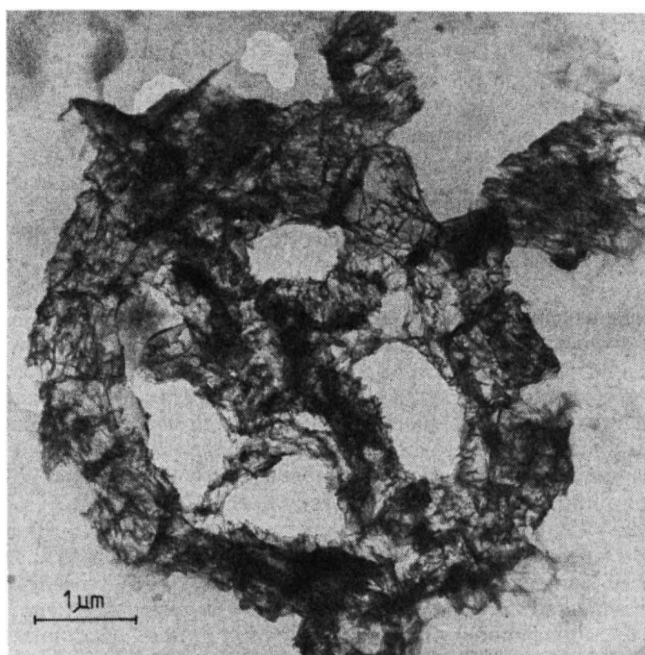


Fig. 2 Decalcified Cretaceous coccolith. The calcite was removed by floating the specimen, mounted on an electron microscope grid, on a 0.005 M EDTA solution for 30 s, then washing in distilled water for 2 h. Slower and more controllable dissolution was achieved by incorporating 5% $\text{NH}_4\text{Cl}/\text{NH}_4\text{OH}$ buffer (pH 10) in the EDTA. Removal of the calcite with a dilute sodium acetate/acetic acid buffer (pH 5) is also possible but tends to rupture the skins. In the example shown the coating on the central 'bridge' feature, often found in coccoliths, has remained intact.

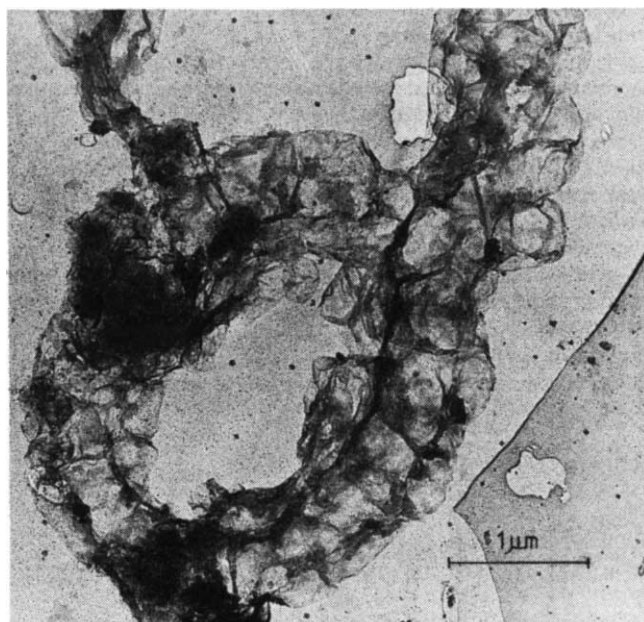


Fig. 3 Decalcified Cretaceous coccolith. The dark area on the left is smectite. The rest of the skin is amorphous.

However, C and H analysis of the chalk before and after treatment with bromine water shows that 68% of the total organic carbon is unavailable for oxidation. This could be because it is incorporated in the coatings as suggested by Hamano and Honjo⁶, possibly bedded between the calcite and the amorphous aluminosilicate layer.

From the present evidence we cannot say whether the inorganic part of the coatings was incorporated during their formation or formed subsequently by preferential precipitation of siliceous material dissolved in the mobile porewater⁸. However this may be, the accurate replicas shown in Figs 2 and 3 could plainly be silicified organic membranes, and demonstrate unequivocally that Cretaceous coccoliths can have a coating closely associated with the calcite.

We thank the Welwyn Hall Research Association for financial support for this work.

Received 25 January; accepted 18 March 1982.

- Hay, W. W. & Mohler, H. P. *J. Paleontol.* **41**, 1505–1541 (1967).
- Pautard, F. G. E. *Biological Calcification: Cellular and Molecular Aspects* (ed. Schraer, H.) 140–141 (North-Holland, Amsterdam, 1970).
- Klavness, D. thesis, Oslo Univ. (1971).
- de Jong, L. W., Dam, W., Westbroek, P. & Crenshaw, M. A. in *The Mechanisms of Mineralisation in the Invertebrates and Plants* (eds Watabe, N. & Wilbur, K. M.) 135–153 (University of South Carolina Press, 1976).
- Hay, W. W. in *Geology and Archaeology of Northern Cyrenaica, Libya* (ed. Barr, F. T.) 149–157 (Petroleum Exploration Society of Libya, 1968).
- Hamano, M. & Honjo, S. *J. geol. Soc. Japan* **75**, 607–614 (1969).
- Weir, A. H. & Catt, J. A. *Clay Miner.* **6**, 97–110 (1965).
- Jeans, C. V. *Clay Miner.* **7**, 311–329 (1968).
- Suess, E. *Geochim. cosmochim. Acta* **34**, 157–159 (1970).

Flow-field variables trigger landing in flies

Hermann Wagner

Max-Planck-Institut für biologische Kybernetik,
Spemannstrasse 38, D-7400 Tübingen 1, FRG

The eyes of moving animals experience a continuously changing flow of stimulation. Information contained in this 'optical flow field' can be extracted to control movement^{1–4}. I present here evidence that pre-landing deceleration in flies is determined by such information: onset of deceleration is triggered when the ratio of the image expansion of a target on the retina to the image size ('relative retinal expansion velocity', RREV) reaches a critical value.

Female houseflies (*Musca domestica* L.) flying freely in a cage (90×70×100 cm) were simultaneously filmed from two sides at 50, 80 or 100 frames per s while landing on black stationary spheres (radius, $r = 0.25$ – 1.0 cm). Frame-by-frame analysis allowed the reconstruction of three-dimensional landing trajectories. A sample stereogram is shown in Fig. 1. I evaluated certain variables, including (1) the fly's velocity (v), defined as the distance travelled along the flight path between two frames, divided by the sampling time Δt ; (2) the distance from the fly to the centre of the target (x); (3) the change of distance with time ($\Delta x/\Delta t$, which is the component of the velocity in the direction of the target); (4) the solid angle subtended by the target on the retina (retinal size a); and (5) the target's retinal expansion velocity ($\Delta a/\Delta t$). As the retinal image is the only source of optical information directly available to the animal, the physical variables (1–3) can only be derived indirectly via the optical variables (4, 5). x , r and a are related as follows⁵:

$$a = 2\pi(1 - \sqrt{1 - (r/x)^2}) \approx \pi r^2/x^2 \quad (\text{for } x \gg r) \quad (1)$$

(if $x = 2r$, the error of the approximation is 7%) and in analogy to Lee¹⁵ for the approximation:

$$\text{RREV} = \frac{1}{a} \frac{da}{dt} = -2 \frac{1}{x} \frac{dx}{dt} \quad \text{or} \quad \text{RREV} = \frac{d}{dt} \ln a = -2 \frac{d}{dt} \ln x \quad (2)$$

Flies reduce their flight velocity before landing and the onset of the final deceleration before landing is taken to define the beginning of the landing phase. Figure 2a shows the velocity profile of different landing trajectories compared by setting zero time to the onset of deceleration. Tests were carried out to determine whether the onset is related to the variables x , a , $\Delta a/\Delta t$, $1/a \times \Delta a/\Delta t$ and v/x . Table 1 shows data obtained from



Fig. 1 Example of a fly trajectory viewed from above. It should be viewed with normal stereoglasses. Open circles and lines represent the head and long axis of the fly, respectively. The solid circle indicates the position and size of the target. The legs could not be seen on the films. Numbers represent the time (s) with onset of deceleration set to zero time; 100 frames per s.

31 landing trajectories, pooled at corresponding times to allow calculation of the coefficients of variation (c.v.) for each variable. As the c.v. is a measure of dispersion defined by standard deviation/mean, variables showing a high c.v. cannot be regarded as critical for the initiation of pre-landing deceleration. At each sampling point the c.v. of both the RREV and the velocity/distance quotient are much smaller than all others, which suggests the prime importance of these variables in initiating onset of deceleration. As velocity and distance are only indirectly accessible, it is concluded that the critical cue is the RREV and not v/x .

The RREV is given by the ratio of retinal expansion velocity to retinal size (equation 2). By correlating these variables it

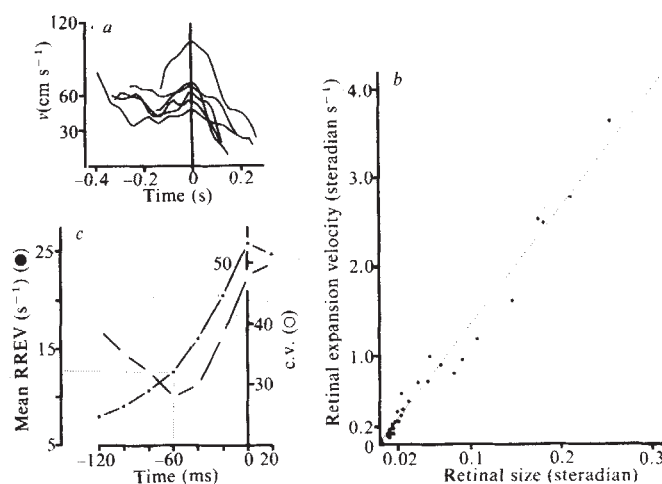


Fig. 2 *a*, Time course of velocity in several examples before landing; the data were evaluated at 50 frames s^{-1} and smoothed with a filter: $v(t) = 0.25(v(t - \Delta t) + 2v(t) + v(t + \Delta t))$. The beginning of the landing phase is defined as the onset of final deceleration before landing (and is set to zero time for all trajectories). As the scatter of deceleration due to digitizing is 100 cm s^{-2} , the measured deceleration must be $< -100 \text{ cm s}^{-2}$ and must not exceed 100 cm s^{-2} before landing. *b*, Dependence of retinal expansion velocity on retinal size of a target for 31 landings 60 ms before onset of deceleration. Spearman's rank correlation coefficient, $\rho = 0.98$. The dependence at other instants is similar. The slope of the regression line (dotted line), which is equal to the RREV, is 13.2 s^{-1} ; the high correlation coefficient demonstrates that the variance of the RREV is low. *c*, Mean (●) and c.v. (○) of the RREV at different time shifts with respect to onset of deceleration (zero time). Coefficient of variation data for the RREV are the same as in Table 1. The c.v. minimum at $t = -60 \text{ ms}$ may indicate the reaction time from the threshold RREV to the onset of deceleration. Dotted lines show the value of the RREV corresponding to -60 ms .

Table 1 Coefficients of variation for different variables with respect to onset of final deceleration

Time (ms)	+20	0	-20	-40	-60	-80	-100	-120
Distance (x)	78	72	66	62	58	55	53	52
Retinal size (a)	128	124	128	127	128	128	137	143
Retinal expansion velocity ($\Delta a/\Delta t$)	143	132	133	128	128	119	113	109
RREV ($1/a \times \Delta a/\Delta t$)	50	48	38	30	28	32	35	39
Velocity/distance (v/x)	48	36	29	27	28	30	32	32

At each instant only one value per trajectory was used to calculate the c.v.. The c.v. of the RREV and the quotient of velocity and distance from fly to target should be identical if the flies approach the target in a straight line [equation (2)]. The differences between these c.v. indicate that this condition is often not fulfilled.

was shown that during the last 120 ms before onset of the final deceleration their dependence on each other was strong, that is, the variance of the RREV was small, confirming the conclusion drawn from Table 1. This is shown in Fig. 2b at 60 ms before onset of deceleration, the instant of maximum dependence corresponding to the minimum of the c.v. of the RREV (Fig. 2c, open circles). The mean value of the RREV increases steadily (Fig. 2c, solid circles). Together with the correlation analysis, this finding leads to the conclusion that, for onset of deceleration to occur, the RREV must exceed a critical value. The minimum of the c.v. at -60 ms (Fig. 2c, open circles) might represent the response delay (see ref. 6). Thus, the threshold value of the RREV is $\sim 13 \text{ s}^{-1}$ (dotted lines in Fig. 2c).

Previous experiments on landing using tethered flying insects⁷⁻¹³ did not yield the same conclusions as presented here from free-flight experiments. In the earlier experiments, extension of the legs before landing ('landing response'), which could not be seen on my films, was taken as a criterion for the initiation of the landing procedure. The conclusions from these experiments are consistent neither amongst themselves nor—partly—with the findings presented here. In particular, a critical retinal target size alone, as proposed by Eckert and Hamdorf¹⁰, does not initiate landing of free-flying flies as shown by the high c.v. (Table 1). Furthermore, as the RREV is almost independent of real target size (equations 1 and 2), the present conclusion can better explain landings on targets of various dimensions than hypotheses which take into account retinal size or retinal expansion velocity alone. Other authors^{8,11-13} found that the landing response could be elicited by centrifugal movement of striped patterns. This simulates target expansion and these results are thus qualitatively consistent with those reported here. However, the dependence of the landing response on the RREV has not yet been investigated, either in experiments with tethered flying flies or in electrophysiological investigations of units with selective response to an expanding stimulus¹⁴.

Formally, the RREV is the inverse of 'time to collision', that is, the time needed to reach a target directly approached at constant velocity¹⁵. As many landing trajectories are neither direct nor of constant velocity, a simple interpretation of the RREV as time to collision is not applicable. The RREV, however, represents the more general case: it gives information about the distance to a target not in absolute, but in relative units dependent on both the translation velocity and the direction of flight with respect to a target. This means that a distance is tagged with a time. Thus, the important cues for landing are times and not distances. Similar arguments might hold for escape responses and even control of cruising flights. The present conclusions agree well with a theory established by Lee and co-workers^{4,15} for humans and birds, and prove that exploiting variables in the optical flow field can provide simple solutions for rather complex orientation tasks.

I thank Professor W. Reichardt for discussions and helpful criticism. Drs J. Brede, H. Bülthoff, E. Buchner, T. Collett, M. Egelhaaf, K. Hausen, T. Poggio, K. Vogt, Chr. Wehrhahn and J. Zeil critically read an early draft of the paper and made many useful suggestions. I thank Dr R. Cook for correcting the English. Computer programs were modified from a program package developed by H. Bülthoff, Chr. Koch and T. Poggio.

Received 21 December 1981; accepted 26 March 1982.

1. Wehner, R. in *Handbook of Sensory Physiology* Vol. 7/6C (ed. Autrum, H.) 287-616 (Springer, Berlin, 1981).
2. Reichardt, W. & Poggio, T. *Q. Rev. Biophys.* **9**, 311-346 (1976).
3. Collett, T. S. *J. comp. Physiol.* **138**, 271-282 (1980).
4. Lee, D. N. & Reddish, P. E. *Nature* **293**, 293-294 (1981).
5. Wagner, H. *Diplomarbeit der Universität Tübingen* (1980).
6. Land, M. F. & Collett, T. S. *J. comp. Physiol.* **89**, 331-357 (1974).
7. Goodman, L. J. *J. exp. Biol.* **37**, 854-878 (1960).
8. Cogshall, J. C. *J. exp. Biol.* **57**, 401-413 (1972).
9. Eckert, H. *Biol. Cybern.* **37**, 235-247 (1980).
10. Eckert, H. & Hamdorf, K. *J. comp. Physiol.* **138**, 253-264 (1980).
11. Perez de Talens, A. F. & Taddei-Ferretti, C. *J. exp. Biol.* **52**, 233-256 (1970).
12. Taddei-Ferretti, C. & Perez de Talens, A. F. *Z. Naturf.* **28e**, 568-592 (1973).
13. Wehrhahn, C., Hausen, K. & Zanker, J. *Biol. Cybern.* **41**, 91-99 (1981).
14. Chillemi, S. & Taddei-Ferretti, C. *J. exp. Biol.* **94**, 105-118 (1981).
15. Lee, D. N. *Perception* **5**, 437-459 (1976).

A peripheral locus for amphetamine anorexia

Michael G. Tordoff, Jill Hopfenbeck,
Larry L. Butcher & Donald Novin

Department of Psychology and Brain Research Institute, University of California, Los Angeles, California 90024, USA

The anorexic action of amphetamine is thought to be mediated by central dopaminergic and noradrenergic neurotransmitter systems¹⁻⁴. However, some of its sympathomimetic effects can be attenuated by denervation⁵ or by depletion of peripheral noradrenaline^{6,7}. This led us to consider the possibility that a component of amphetamine anorexia might be due to excitation of the sympathetic innervation of the viscera. To test this, the amphetamine dose-response curves of free-feeding rats with visceral sympathetic denervation by coeliac ganglionectomy (GAN) were compared with those of sham-operated controls (SHM). We report here that ganglionectomized rats were protected from the anorexic effects of low doses of amphetamine (0.2 or 0.4 mg per kg interaperitoneal, i.p.). We suggest that the anorexia might result from enhanced hepatic glycogenolysis⁸, and the subsequent stimulation of hepatic metabolic receptors that inhibit feeding⁹.

Sixteen male Long-Evans hooded rats (Charles River) weighing 216-246g were maintained on a 12:12 h light/dark cycle (dark period beginning at 17.00) with food (Wayne lab blox, Chicago) and water continuously available. Half the animals were subjected to coeliac ganglionectomy by the method of Lambert¹⁰; the others were subjected to identical procedures except that the ganglion was not excised. After surgery, GAN rats showed a transient 2-day hypophagia accompanied by diarrhoea, which is a good indication of complete ganglionectomy. Rats were given 24 days to recover after surgery. They were then tested in groups of four (two GAN and two SHM) every four days. All received four saline injection trials, to habituate them to the drug administration and measurement regimen, followed by counterbalanced injections of saline or (+)amphetamine sulphate (0.2, 0.4, 0.6, 0.8 or 1.0 mg per kg i.p.).

The procedure used for habituation and testing was as follows: 1 h before the start of the dark period, each rat was

weighed and placed in a cage designed to measure activity by photobeam crossings¹¹. Food was available on the cage floor and tap water could be obtained from a drinking tube. Just before the lights were turned off, each rat was injected and a fresh quantity of food was provided. Food intake was measured after 30, 60, 90 and 120 min by weighing the remaining food and subtracting this value from its initial weight. Activity was monitored continuously but, for purposes of data analysis, was grouped into 30-min intervals.

Separate three-way analyses of variance (group $\times$ dose $\times$ time), followed by Tukey's *post hoc* tests, were done on the food intake and activity measurements. Amphetamine caused a dose-related anorexia ($F(5, 70) = 8.31, P < 0.001$): during the

first 90 min after injection, the food intake of SHM rats was significantly depressed by all doses of amphetamine. However, GAN rats showed no decrease in consumption in response to amphetamine at either 0.2 or 0.4 mg per kg ($F(5, 70) = 2.42, P < 0.05$; see Fig. 1a). During the same interval, both groups showed a comparable dose-related hyperactivity ($F(5, 70) = 28.59, P < 0.001$; see Fig. 1b). Thus, the feeding dose-response curves of SHM and GAN rats differed, but their activity responses were the same. Disrupting the sympathetic innervation of the viscera caused an attenuation of anorexia that was independent of the hyperkinetic action of amphetamine. This sympathetic mechanism accounted for a relatively large proportion of the anorexia at low doses of amphetamine, and relatively little at higher doses.

It is unlikely that nonspecific effects of the coeliac ganglionectomy affected the results. The food intake of SHM and GAN rats after injections of saline was statistically indistinguishable, as was activity after all injections. Furthermore, GAN rats ate more food than controls after low doses of amphetamine, which suggests there was no general debilitation.

The coeliac ganglion innervates most of the abdominal viscera. However, we consider the most likely site for the peripheral anorexic action of amphetamine to be the liver. The liver has considerable sympathetic innervation¹² and is involved in the glucostatic regulation of feeding. Russek⁹ postulated that hepatic receptors sensitive to glucose or one of its metabolites respond to the liberation of glucose from glycogen, so that promotion of glycogenolysis stimulates hepatic receptors, which inhibit feeding. In support of this hypothesis, glucose, 2-deoxy-D-glucose and adrenaline have greater effects, with shorter latency, when administered intraportally or intraperitoneally than systemically^{8,13,14}. In addition, vagotomy attenuates the anorexia induced by glucose, glucagon, adrenaline and amphetamine¹⁴⁻¹⁷. Both electrical stimulation of hepatic sympathetic fibres and i.p. injection of amphetamine induce glycogenolysis^{8,18}. Finally, glucagon and amphetamine are more potent in rats with high rather than low glycogen levels^{8,17}. A possible mechanism for peripherally-induced amphetamine anorexia could therefore involve the stimulation of sympathetic hepatic nerve fibres, hence inducing glycogenolysis and subsequent activation of metabolic receptors capable of inhibiting feeding^{19,20}.

Because the coeliac ganglion has both efferent and afferent hepatic innervation, including a large component of the vagus nerve, this study did not determine whether amphetamine acts either directly on the liver or its local innervation, or indirectly via a sympathetic discharge which originates in the CNS and which activates sympathetic fibres that travel to the liver. However, recent evidence suggests that direct stimulation of the liver by amphetamine may be relatively unimportant, as amphetamine-induced anorexia was more effectively blocked by coeliac ganglionectomy than by subdiaphragmatic vagotomy but adrenaline-induced anorexia was attenuated to the same extent by these two surgical procedures¹⁶. As adrenaline is a sympathomimetic which, unlike amphetamine, does not easily enter the brain, and because the coeliac ganglion provides hepatic sympathetic innervation, the additional protection from amphetamine anorexia afforded by coeliac ganglionectomy may reflect interruption of sympathetic discharge of CNS origin.

Blundell and Burridge²¹ suggested that amphetamine anorexia consists of additive noradrenergic and dopaminergic components. In their model, the anorexia attributable to noradrenaline (NA) increases over low doses of amphetamine but reaches a maximum ~ 0.5 mg per kg. The anorexia attributable to dopamine (DA) has little effect in this range, but increases rapidly for doses > 0.5 mg per kg. Although there is evidence which seems to contradict the Blundell and Burridge model²²⁻²⁴, it is tempting to speculate that the dual anatomical (brain and periphery) and neurochemical (DA and NA) mechanisms may correspond. One might represent a centrally-mediated dopaminergic factor, and the other, a peripherally-mediated noradrenergic sympathetic discharge factor. Our

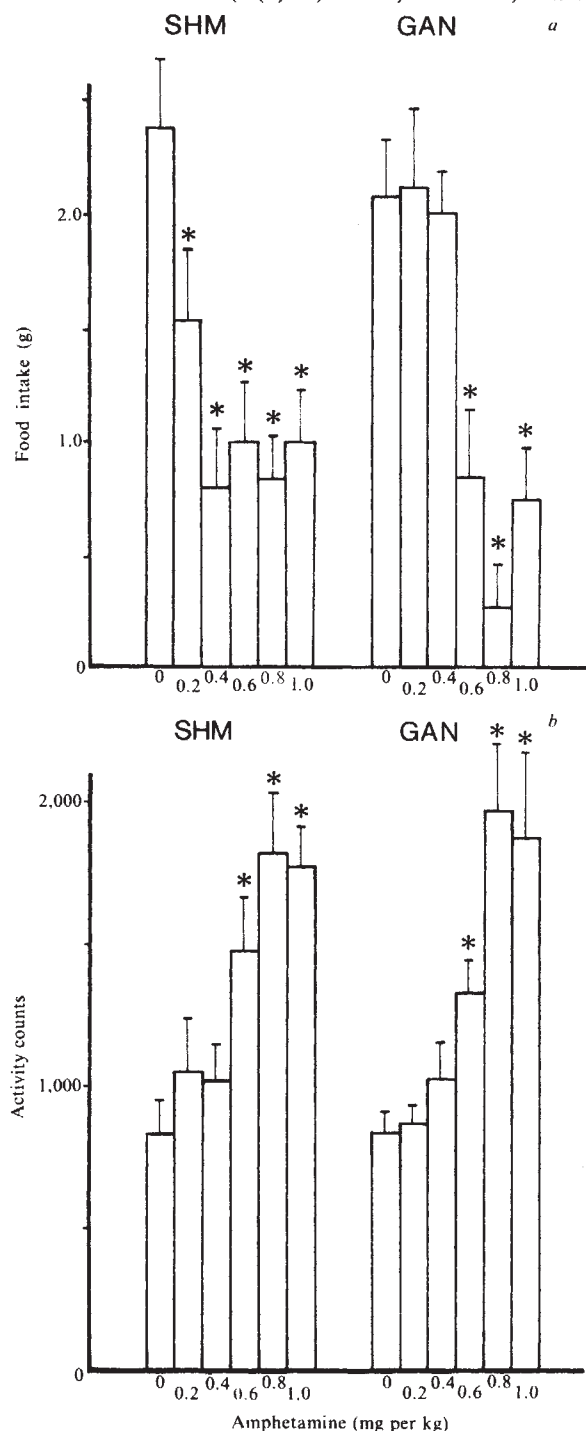


Fig. 1 The 90-min cumulative food intake (a) and 90-min cumulative activity (b) of sham (SHM) and coeliac ganglionectomized (GAN) rats were measured after administration of various doses of amphetamine. * $P < 0.05$ compared with saline-injected controls (Tukey's test).

finding that GAN and SHM animals showed differences only with doses of amphetamine <0.5 mg per kg provides support for such a speculation, as the Blundell and Burridge model predicts this to be the dose where the major contribution to amphetamine anorexia changes from NA to DA.

Previous reports have negated the possibility of a peripheral locus for the anorexia induced by amphetamine, but these studies have used large doses, food-deprived animals or insensitive measuring techniques^{2,3}. We have found that these conditions are the reverse of those required to observe differences in feeding between sympathetically denervated subjects and their controls. This suggests that the usual techniques for screening anorexic drugs using large doses and food-deprived rats may be inappropriate, because such tests would not recognize active anorexigens that had sympathetico-hepatic effects.

This research was supported by National Institute of Health NS07687 to D. Novin and NS10928 to L.L.B.

Received 20 November 1981; accepted 18 March 1982.

1. Cole, S. O. *Psychol. Bull.* **79**, 13–20 (1973).
2. Harris, S. C., Ivy, A. C. & Searle, L. M. *J. Am. med. Ass.* **134**, 1468–1475 (1947).
3. Nathanson, M. H. *J. Am. med. Ass.* **108**, 528–531 (1937).
4. Leibowitz, S. F. *Brain Res.* **98**, 529–545 (1975).
5. Lockett, M. F. *Br. J. Pharmac.* **5**, 485–496 (1950).
6. Burn, J. H. & Rand, M. J. *J. Physiol., Lond.* **144**, 314–336 (1958).
7. Muscholl, E. A. *Rev. Pharmac.* **6**, 107–128 (1966).
8. Russek, M. & Stevenson, J. A. *F. Physiol. Behav.* **8**, 245–249 (1972).
9. Russek, M. *Nature* **197**, 79–80 (1963).
10. Lambert, R. *Surgery of the Digestive System in the Rat*, 484 (Charles Thomas, Springfield, Illinois, 1965).
11. Hodge, G. K. & Butcher, L. L. *Pharmac. Biochem. Behav.* **10**, 695–705 (1979).
12. Williams, P. L. & Warwick, R. *Functional Neuroanatomy of Man*, 1068–1083 (Churchill-Livingstone, Edinburgh, 1975).
13. Russek, M., Lora-Vilchis, M. C. & Islas-Chaires, M. *Physiol. Behav.* **24**, 157–161 (1972).
14. Novin, D., Vanderweele, D. A. & Rezek, M. *Science* **181**, 858–860 (1973).
15. Novin, D., Sanderson, J. & Gonzalez, M. *Physiol. Behav.* **22**, 107–113 (1979).
16. Tordoff, M. G., Novin, D. & Russek, M. *J. comp. physiol. Psychol.* **96**, 361–375 (1982).
17. Vanderweele, D. A., Geiselman, P. J. & Novin, D. *Physiol. Behav.* **23**, 155–158 (1979).
18. Shimazu, T. & Fukuda, A. *Science* **150**, 1607–1608 (1965).
19. Mogenson, G. J. *Physiol. Behav.* **3**, 133–136 (1968).
20. Russek, M. & Bruni, E. *Ann. Inst. Univ. Auton. Mexico*, **41**, Ser. Biol. Exp. **1**, 17–24 (1970).
21. Blundell, J. E. & Burridge, S. L. in *The Treatment of Obesity* (ed. Munro, J. F.) 53–83 (University Park Press, Baltimore, 1979).
22. Franklin, K. B. J. & Herberg, L. J. *Neuropharmacology* **16**, 45–46 (1977).
23. Frey, H.-H. & Schulz, R. *Biochem. Pharmac.* **22**, 3041–3049 (1973).
24. Schmitt, H. J. *Pharmac. Paris* **4**, 285–294 (1973).

Effect of chloroform on charge movement in the nerve membrane

J. M. Fernández*, F. Bezanilla* & R. E. Taylor*†

* Department of Physiology, Ahmanson Laboratory of Neurobiology and Jerry Lewis Neuromuscular Research Center, University of California, Los Angeles, California 90024, USA and The Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA
† Laboratory of Biophysics, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20205, USA and The Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

It is generally accepted that membrane-bound charges or dipoles moving under the influence of the transmembrane electric field produce conformational changes that result in the opening and closing of ion-selective channels¹. The location of these charged groups or dipoles is unknown but it has been speculated that they are sited either inside the channel macromolecule or embedded in the lipid bilayer². We report here that in the squid axon membrane, the general anaesthetic chloroform reduces the magnitude of the gating currents but has no effect on their kinetics. In contrast, chloroform increases the rate of charge translocation of the lipophilic ion dipicrylamine (DPA⁻). These results indicate that the gating charges responsible for the opening and closing of the Na⁺ channel do not move in the bulk lipid of the membrane as do DPA⁻ molecules, but that they move in a physical domain where their kinetics are unaffected by changes in the structural parameters of the lipid matrix. It is therefore unlikely that general anaesthetics act to modify the kinetics of the opening and closing of ionic channels via modifications of the structural parameters of the lipid matrix³.

Experiments were performed on internally perfused and voltage-clamped squid giant axons using the methods of Bezanilla *et al.*⁴ to record Na⁺ and gating currents. Figure 1a shows that perfusion of the axon with a solution containing 60 mM chloroform reduced dramatically the magnitude of the inward Na⁺ current. To observe sodium and gating currents simultaneously, the axon was bathed in a reduced Na⁺ concentration (Fig. 1b, trace 1). In these conditions, 43 mM of chloroform added to the internal perfusion solution reduced the Na⁺ current in parallel with the gating current (Fig. 1b, trace 2). Both effects were reversed on removing the chloroform (Fig. 1b, trace 3). The effect of chloroform on the Na⁺ and gating currents was not reversed by hyperpolarizing prepulses, suggesting that these effects are not related to the mechanisms of inactivation.

To analyse the effects of chloroform on gating currents, the Na⁺ current was eliminated by substituting the external Na⁺ with the impermeant ion Tris and by the addition of 300 nM tetrodotoxin (TTX). Figure 1b, traces 4, 5, shows that

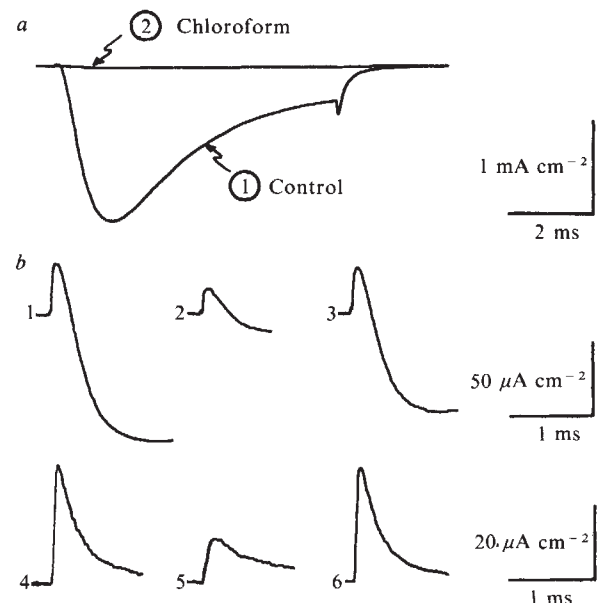


Fig. 1 Effect of chloroform on the Na⁺ inward current and the gating currents. *a*, Trace 1, inward Na⁺ current before treatment with chloroform; trace 2, abolition of the Na⁺ current when the axon was internally perfused with a solution containing 60 mM chloroform. Traces 1 and 2 were obtained under voltage clamp by pulsing to -10 mV from a holding potential (HP) of -70 mV. Leakage and linear capacitive currents were subtracted using the P/4 technique with a subtracting holding potential (SHP) of -140 mV (for details of the technique see ref. 4). The external solution was 440 mM NaCl, 50 mM MgCl₂, 10 mM CaCl₂, pH 7.2. The internal solution was 100 mM *N*-methylglucamine(NMG) glutamate, 100 mM NMG fluoride and ~ 400 mM sucrose adjusted to make the internal and external solution isosmotic, pH 7.2, 5°C. *b*, Trace 1, axon with low external Na⁺ allowed measurement of the ON gating current as well as the Na⁺ inward current. Trace 2, when perfused with a solution containing 43 mM chloroform, a significant reduction in the Na⁺ current was paralleled by a reduction in the gating current. Trace 3, on removal of chloroform, both effects were reversed. Trace 4, ON gating currents measured in the same axon as in traces 1–3, after eliminating the Na⁺ current with an external solution containing Tris and 300 nM TTX. Trace 5, partial decrease in the size of the ON gating current when the axon was perfused with the same solution as in trace 2. Trace 6, after removing chloroform, the gating current was almost completely recovered. The full effect of chloroform on the Na⁺ and gating currents occurred within seconds of beginning the perfusion with chloroform-containing solutions; recovery was complete within <1 min after removal of chloroform. Traces 1–6 were obtained by pulsing to 30 mV from HP = -70 mV using the P/4 procedure with a SHP of -140 mV. The external solution (traces 1–3) was 88 mM NaCl, 360 mM Trizma 7.2, 50 mM MgCl₂, 10 mM CaCl₂. Traces 4–6 were obtained using an external solution containing 450 mM Trizma 7.2, 50 mM MgCl₂, 10 mM CaCl₂ and 300 nM TTX. The internal solution in each case was 100 mM NMG fluoride, 100 mM NMG glutamate and ~ 400 mM sucrose, pH 7.2, 5°C. Where indicated, chloroform was added to the internal solution, which was stirred for 30 min in a closed glass cylinder. The chloroform concentrations correspond to the fraction (v/v) added. In the experiments reported here we did not add chloroform to the external solution because the experimental chamber consisted of acrylic plastic (Perspex).

chloroform reduced the size of the gating current. As expected, this effect was reversible (Fig. 1*b*, trace 6).

Gating currents during a depolarizing or hyperpolarizing pulse (ON gating) can be described by the sum of two voltage-dependent exponential relaxations⁵ (Fig. 2*a*, trace 1); the solid line indicates a two-exponential fit with time constants $\tau_F = 126 \mu\text{s}$ and $\tau_S = 571 \mu\text{s}$. When the axon was internally perfused with a solution containing 60 mM chloroform, the remaining gating current (Fig. 2*a*, trace 2) could again be fitted with two exponentials having time constants $\tau_F = 112 \mu\text{s}$ and $\tau_S = 587 \mu\text{s}$ which, within our resolution, are identical to those obtained in the fit of trace 1. However, the amplitude of the fast exponential component was reduced by about half but there was no significant change in the amplitude of the slow exponential component. This effect was seen for both depolarizing and hyperpolarizing pulses. The difference record between the gating current before addition of chloroform (Fig. 2*a*, trace 1) and during chloroform (trace 2), shown in Fig. 2*b*, is fitted by a single exponential (solid trace) with a time constant $\tau_F = 126 \mu\text{s}$. Figure 2*b* represents the fraction of the fast component eliminated by chloroform. A more dramatic reduction of the fast exponential component is shown in Fig. 2*c*.

The main effect of chloroform was to decrease the amplitude of the fast component of gating with no significant effect on the kinetics of either the remaining fast or slow components. These results were qualitatively reproduced in several experiments, with variations only in their magnitude (see Fig. 2*a*, *c*) due to our present inability to control accurately the concentration of chloroform. The magnitude of the effect obtained was dependent on internal and external perfusion rates, therefore the actual concentration of chloroform reaching the membrane was probably significantly smaller than the values given throughout this paper. The concentrations given were estimated from the fraction of chloroform (v/v) added to the internal perfusion solution.

Several models have been proposed for the action of general anaesthetics such as chloroform on the structure of lipid bilayers. These include modifications of such structural parameters as thickness, dielectric constant, dipole potential and the degree of order (for example, fluidity) of the lipid hydrocarbon chains (see refs 3, 6). Lipophilic ions such as DpA^- and tetraphenylboron have proved useful in elucidating these structural parameters of lipid bilayers^{3,7}. The general behaviour of lipophilic ions in membranes is phenomenologically similar to that of gating currents, because they bear a charge but are trapped inside the membrane⁷. It has been shown that changes in the structure and composition of the bilayer greatly affect the transport of these lipophilic ions. For example, Reyes and Latorre³ demonstrated that pharmacological concentrations of the neutral general anaesthetics chloroform and benzyl alcohol significantly increase the translocation rate of the hydrophobic ion tetraphenylborate in lipid bilayers, indicating that these anaesthetics change the dielectric constant of the lipid matrix. As these physical changes should affect the movement of any membrane-bound charge, a similar effect would be expected for the charge movement of the gating particles responsible for the opening and closing of Na^+ channels. However, the results presented here show that this is not the case. The gating current kinetics of the fast and slow components were unmodified by chloroform, which indicates that the mechanism of action of chloroform on gating currents is very different from that on the charge movement induced by lipophilic ions.

To confirm that chloroform has the same general effect on the membrane of squid giant axon as it does on lipid bilayers, we incorporated the lipophilic ion DpA^- into the squid giant axon membrane and measured the effect of chloroform on the induced charge movement. As shown in Fig. 3*a*, DpA^- produced a very large charge movement having characteristics similar to those described by Benz and Conti⁸ (details of these results will be published elsewhere). When the axon was

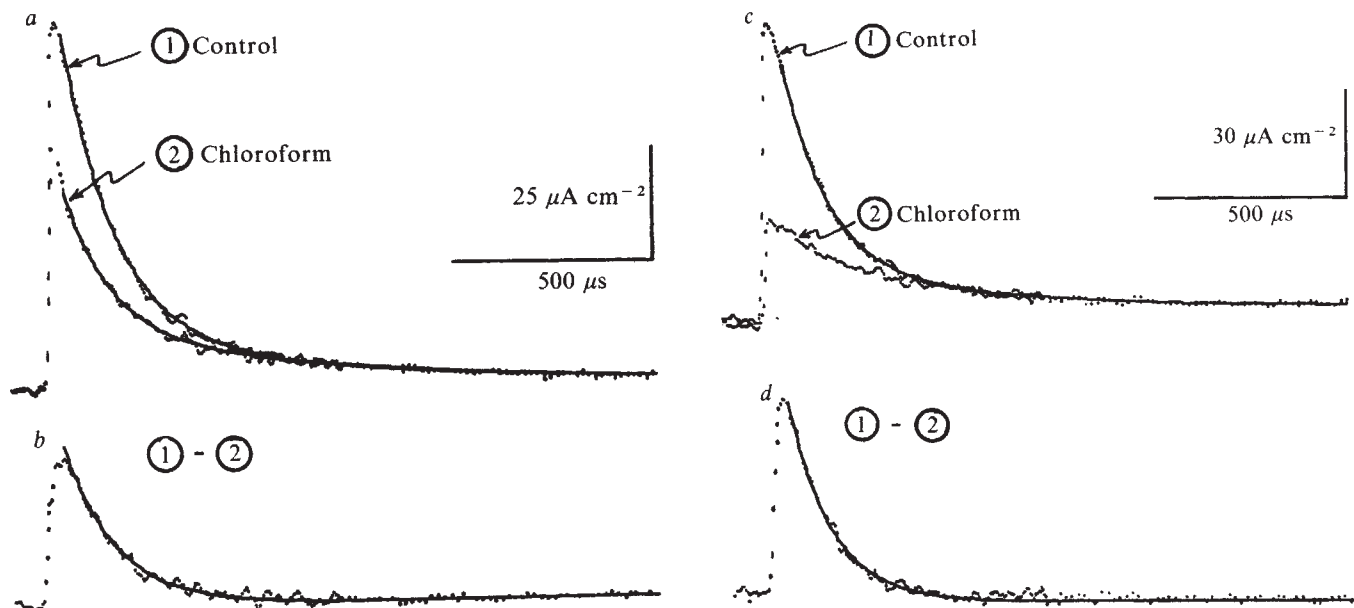


Fig. 2 Effect of chloroform of the ON gating currents. *a*, Trace 1 shows an ON gating current measured using the P/4 procedure in the absence of ionic currents for a test pulse of +100 mV from a holding potential of -70 mV (SHP = -150 mV). The solid line represents a two-exponential fit with a fast time constant, $\tau_F = 126 \mu\text{s}$ and a slow time constant, $\tau_S = 571 \mu\text{s}$. Trace 2 shows the effect of internally perfusing the axon with a solution containing 60 mM chloroform; this trace could again be fitted with two exponentials having time constants $\tau_F = 112 \mu\text{s}$ and $\tau_S = 587 \mu\text{s}$. Chloroform reduced the amplitude of the fast component to about half without altering the kinetics of the remaining fast and slow components. *b*, Numerical difference between traces 1 and 2 of *a*, which corresponds to the fraction of the fast component eliminated by chloroform. The solid line represents a single exponential fit with time constant $\tau_F = 126 \mu\text{s}$. *c*, A more dramatic reduction of the fast component for a different axon internally perfused with a solution containing 60 mM chloroform. Trace 1 corresponds to the ON gating current for a test pulse of +110 mV from a holding potential of -70 mV (SHP = -140 mV). A two-exponential fit (solid line) gave the time constants $\tau_F = 108 \mu\text{s}$ and $\tau_S = 508 \mu\text{s}$. Trace 2 shows the gating current remaining when the axon was perfused with the chloroform-containing solution; in this case the fast component of trace 1 was reduced by ~85%. *d*, The numerical difference between traces 1 and 2 in *c*; the solid line corresponds to a single exponential fit with a time constant $\tau_F = 111 \mu\text{s}$.

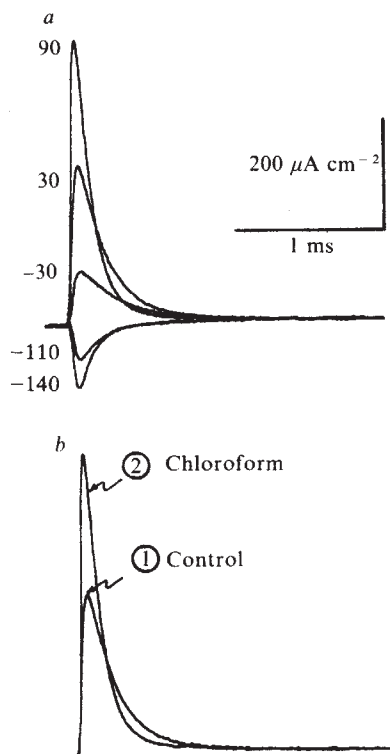


Fig. 3 Charge movement induced by the lipophilic ion dipicrylamine in the absence of ionic currents measured using the P/4 procedure. *a*, Charge movement measured after perfusing the axon internally and externally with solutions containing 10^{-5} M DpA⁻ for ~5 min. The currents correspond to pulses applied, from a holding potential of -70 mV, to the indicated membrane potential. As the induced charge movement was so large, correction for the normal gating charge movement was unnecessary. *b*, Effect of internally perfusing the axon with a solution containing 43 mM chloroform on the kinetics of the DpA⁻-induced charge movement. In contrast to the effect observed for the ON gating current, the rate of DpA⁻-induced charge movement was increased by chloroform (traces 1 and 2 were obtained by pulsing to +30 mV from a HP of -70 mV with a SHP of -140 mV). The effect was reversed after removal of chloroform. The data shown are for the same axon and solutions as used for the axon of Fig. 1*b* (traces 4–6), and demonstrate that the contrasting effects of chloroform on the Na⁺ gating currents and the DpA⁻-induced charge movement can be observed in the same axon in identical conditions.

perfused with a solution containing 43 mM chloroform, the translocation rate of DpA⁻ was increased significantly as indicated by a larger initial current and a faster decay (Fig. 3*b*). The voltage dependence of the DpA⁻ charge movement was unmodified by chloroform, indicating that the observed change in the translocation rate represented a reduction in the activation energy that the DpA⁻ molecules must overcome to cross the membrane, in good agreement with the observations of Reyes and Latorre³ for artificial bilayers. Thus chloroform has similar effects on the structural parameters of squid axon and lipid bilayer membranes.

These experiments therefore indicate that Na⁺ channel gating charges do not diffuse through the bulk lipid matrix of the membrane as the DpA⁻ molecules do. It is therefore questionable to use hydrophobic ions as analogues of the Na⁺ gating charge movement. This conclusion has important implications as it renders unlikely those classes of model in which the gating charges of the Na⁺ channel move freely in the lipid matrix and in which gating is due to the aggregation of several charged proteins that are oriented by the field to form an operational channel². These experiments also rule out models in which the action of general anaesthetics is to enhance the rate at which the Na⁺ gating charges move between closed and open states,

subsequently altering excitability³. The results suggest that chloroform decreases the number of operational channels by preventing the movement of the charge that produces the fast component of gating current.

The major difficulty in the study of gating currents has been the inability to examine the various components separately. Our experiments suggest the possibility of using chloroform as a tool to dissociate the fast from the slow component and to study the correlation between these components and the Na⁺ conductance to elucidate further the gating processes of voltage-dependent channels.

We thank Drs S. Krasne, J. Vergara and M. White for helpful comments on the manuscript. This work was supported by a Grass Fellowship to J.M.F. and by USPHS grant GM 30376.

Received 4 December 1981; accepted 16 March 1982.

1. Armstrong, C. M. & Bezanilla, F. *Nature* **242**, 459–461 (1973).
2. Mueller, P. *Ann. N.Y. Acad. Sci.* **264**, 247–263 (1976).
3. Reyes, J. & Latorre, R. *Biophys. J.* **28**, 259–280 (1979).
4. Bezanilla, F., Taylor, R. E. & Fernández, J. M. *J. gen. Physiol.* **79**, 21–40 (1982).
5. Bezanilla, F. & Taylor, R. E. *Biophys. J.* **23**, 479–484 (1979).
6. Franks, N. P. & Lieb, W. R. *J. molec. Biol.* **133**, 469–500 (1979).
7. Szabo, G. *Ann. N.Y. Acad. Sci.* **303**, 266–280 (1977).
8. Benz, R. & Conti, F. *J. Membrane Biol.* **59**, 91–104 (1981).

Regulation of muscarinic ligand binding sites by nerve growth factor in PC12 pheochromocytoma cells

J. E. Jumblatt* & A. S. Tischler†

* Department of Biochemistry and Pharmacology and

† Department of Pathology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA

The rat pheochromocytoma cell line PC12 (ref. 1) possesses many characteristics of adrenal chromaffin cells, including the ability to synthesize, store and release catecholamines in response to acetylcholine and other depolarizing agents^{2–4}. When exposed to 2.5S nerve growth factor (NGF), PC12 cells cease to divide, and exhibit a number of neurone-like characteristics, including the extension of neurites, development of voltage-sensitive ion channels and electrical excitability, and increased capacity to synthesize and release acetylcholine^{3–8}. Biochemical and pharmacological evidence indicates that both adrenal chromaffin cells and sympathetic neurones contain, in addition to ganglionic-type nicotinic receptors, muscarinic cholinergic receptors proposed to modulate catecholamine release^{9–13}. Catecholamine release from PC12 cells is directly stimulated by nicotinic, but not muscarinic, cholinergic agonists³. Although ion flux through acetylcholine-activated channels in PC12 cells is inhibited by high concentrations of the potent muscarinic antagonist quinuclidinyl benzilate (QNB)¹⁴, the possibility that these cells may contain functional muscarinic receptors has been largely overlooked. We show here that PC12 cells contain saturable, high-affinity binding sites for ³H-(–)-quinuclidinyl benzilate (³H-QNB), which seem pharmacologically comparable with muscarinic receptors described in rodent brain^{15,16}, smooth muscle^{17,18} and adrenal medulla¹². Moreover, treatment with NGF produces a marked elevation of specific ³H-QNB binding. PC12 cells thus may provide a useful system for investigating the biochemical mechanisms, cellular distribution and regulation of muscarinic receptors by hormones and trophic signals.

PC12 cells (passages 23–32) were derived from the clone originally isolated by Greene and Tischler¹. Cells were cultured in collagen-coated 10-cm polystyrene dishes as described previously^{1,5}. To treat cells with NGF for binding experiments, replicate cultures were inoculated at a density of 4×10^4 cells per cm² (except where otherwise indicated), and cultured for

12–16 days in control medium or in medium supplemented with 50 ng ml⁻¹ 2.5S NGF (ref. 19). The medium was changed every third day. To collect cells, the culture medium was replaced by 5 ml Ca²⁺, Mg²⁺-free phosphate-buffered saline²⁰ (CMF-PBS) containing 1 mM EDTA, and cells were gently removed using a rubber policeman. Complete detachment of cell bodies and neurites was monitored by phase microscopy. Cells were washed twice by centrifugation (1,000g, 10 min) and resuspension in 0.1 vol. of 320 mM sucrose, 50 mM phosphate, 0.1 mM EDTA pH 7.2 (SPE buffer) at 4 °C, and counted using a haemocytometer. To prepare membranes, cells were homogenized for 30 s with a Potter-Elvehjem homogenizer and centrifuged for 10 min at 1,000g to pellet the nuclei. Homogenization and centrifugation were repeated on the nuclear pellet, and the combined supernatants were centrifuged at 120,000g for 60 min to obtain a crude membrane (mitochondria + microsomes) fraction. The membrane pellet was resuspended in SPE buffer (3–5 mg protein per ml) and used within 12 h for the binding measurements. All fractionation steps were performed at 0–4 °C.

The concentration dependence of ³H-QNB binding at equilibrium to PC12 cell membranes was determined in the absence and presence of 10⁻⁶ M atropine (Fig. 1). Scatchard analysis²¹ of specific ³H-QNB binding (Fig. 1b; total minus nonspecific binding) indicated a single, high-affinity class of binding sites having a $K_d = 0.146 \pm 0.07$ nM, $B_{max} = 32.7 \pm 7.2$ fmol per mg protein. These results are in close agreement with the reported affinity of QNB binding to muscarinic cholinergic receptors in rat brain¹⁵, smooth muscle^{17,18} and adrenal medulla^{11,12}. Interestingly, the density of QNB binding sites in PC12 cell membranes is comparable with that reported for rat adrenal medulla membranes¹², but more than 10-fold higher than in bovine adrenal medulla¹¹. Assuming 100% recovery of receptor-containing membranes, PC12 cells have ~1,400 specific QNB binding sites per cell. By comparison, Patrick and Stallcup²² reported these cells to contain 5,000 binding sites per cell for

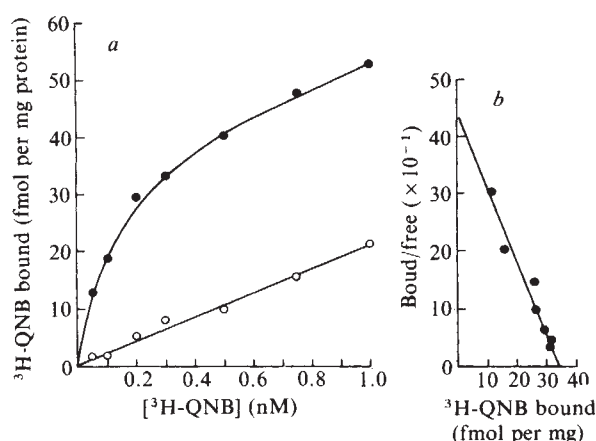


Fig. 1 Binding of ³H-QNB to membranes from untreated PC12 cells as a function of ³H-QNB concentration. PC12 cell membranes (100–160 µg protein) were preincubated for 10 min at 30 °C in microfuge tubes containing 0.9 ml of 50 mM Na⁺+K⁺ phosphate buffer pH 7.2 plus or minus atropine sulphate (10⁻⁶ M final concentration). A volume of ³H-QNB (NEN; 31.2 Ci mmol⁻¹) was then added to each tube to give the indicated final ³H-QNB concentrations. Following incubation at 30 °C for 60 min, 0.8-ml aliquots were removed from each sample and filtered under suction through 25 mM Whatman GF/F glass fibre filters. Each filter was washed four times with 5 ml of ice-cold 50 mM phosphate, then dried and counted in 10 ml Beckman Ready Solv HP scintillation fluid using a Packard Tri-Carb liquid scintillation counter. All points are the means of duplicate determinations, which deviated by <15% from the indicated values. The protein content was determined by the Lowry method²⁴. *a*, Total binding of ³H-QNB (●); nonspecific binding in the presence of 10⁻⁶ M atropine (○). *b*, Scatchard plot of specific ³H-QNB binding from the experimental data shown in *a*.

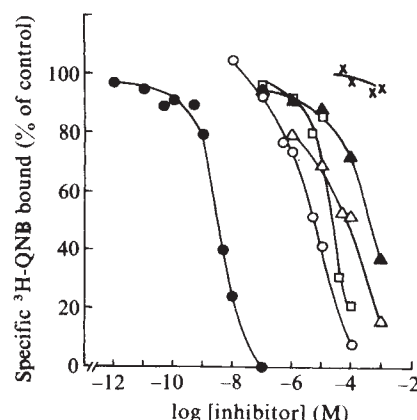


Fig. 2 Competition for specific ³H-QNB binding to PC12 membranes by different concentrations of cholinergic drugs: atropine (●), oxotremorine (○), pilocarpine (□), (+)tubocurarine (▲), carbamylcholine (△), nicotine (×). Binding assays were performed essentially as described in Fig. 1 legend except that different concentrations of competing ligand were present during the initial preincubation and a single concentration of ³H-QNB (0.2–0.6 nM) was present during the final incubation. Nonspecific binding has been subtracted from total binding and per cent of maximum specific binding in controls without inhibitor has been plotted. All points represent the means of duplicate or triplicate measurements, which ranged within 15% of the means.

¹²⁵I- α -bungarotoxin—a selective probe of nicotinic cholinergic receptors at neuromuscular junctions. The relationship of α -bungarotoxin binding sites to functional nicotinic receptors in PC12 cells is unclear, however, because the toxin fails to block the nicotinic receptor-mediated responses in these cells²².

To evaluate the pharmacological specificity of ³H-QNB binding, various cholinergic agents were tested at several concentrations for inhibition of specific QNB binding at a fixed concentration of radioligand (Fig. 2). Corrected IC₅₀ values obtained from these data are listed in Table 1. The order of potency (atropine > oxotremorine > pilocarpine > carbamylcholine > (+)tubocurarine > nicotine) and the IC₅₀ values for these agents correspond closely to the pharmacological selectivity of ligand binding to muscarinic cholinergic receptors in brain and other tissues^{11,15–18}.

Muscarinic receptors exhibit complex binding characteristics for agonists (Hill coefficients < 1), whereas binding of antagonists usually conforms to a simple mass action isotherm (Hill coefficients ~ 1)²³. The basis for this complexity is unknown, but may reflect receptor heterogeneity or conformational shifts²³. Hill analysis (not shown) of the inhibition data in Fig. 2 suggested similar complexity for agonist but not antagonist binding to PC12 membranes. Estimated Hill coefficients were 0.55 for carbamylcholine, 0.74 for oxotremorine, 0.81 for pilocarpine and 1.06 for atropine.

NGF treatment of PC12 cells is known to induce choline acetyltransferase, increased synthesis, storage and release of

Table 1 Relative abilities of cholinergic drugs to inhibit specific ³H-QNB binding to PC12 cell membranes

Drug	IC ₅₀ * (M)
Atropine	1.2 × 10 ⁻⁹
Oxotremorine	1.6 × 10 ⁻⁶
Pilocarpine	7.4 × 10 ⁻⁶
Carbamylcholine	2.6 × 10 ⁻⁵
(+)Tubocurarine	1.5 × 10 ⁻⁴
Nicotine	> 10 ⁻³

Values are derived from the data presented in Fig. 2. IC₅₀* is the concentration of inhibitory drug which reduced specific ³H-QNB binding by 50%, corrected for the shift in radioligand occupancy using the formula $IC_{50}^* = IC_{50} / (1 + [L] / K_d)$ where K_d is the measured equilibrium dissociation constant for ³H-QNB (0.146 nM) and $[L]$ is the concentration of radioligand present (0.2–0.6 nM).

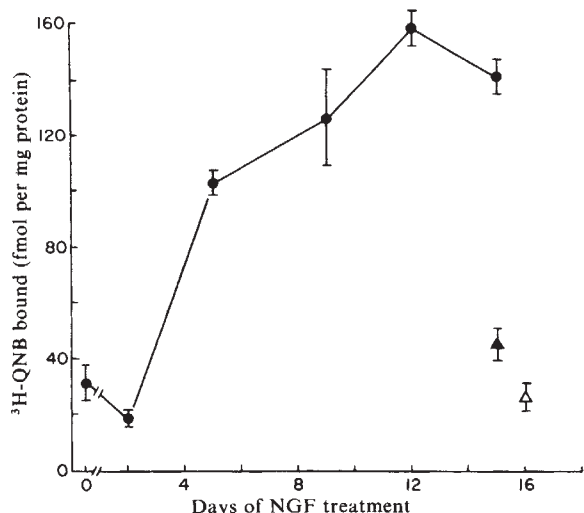


Fig. 3 Effect of NGF treatment on specific ^3H -QNB binding. Parallel cultures were plated and maintained in medium with or without NGF (50 ng ml^{-1}). Medium was changed every third day. Specific binding was determined as described in Fig. 1 legend, using 10^{-9} M ^3H -QNB in the presence and absence of 10^{-6} M atropine. Values represent means $\pm$ s.e.m. of triplicate determinations. NGF-treated cells plated at a density of 4×10^4 cells per cm^2 (●); untreated cultures plated at a density of 4×10^4 cells per cm^2 (▲); untreated cultures plated at 2×10^3 cells per cm^2 (△).

acetylcholine, elevation of acetylcholinesterase activity and the appearance of clustered vesicular structures resembling cholinergic synaptic vesicles^{1,6-8}. When we investigated the effects of NGF, on QNB binding, we found that the level of specific ^3H -QNB binding sites in PC12 membranes increased markedly during 15 days of NGF treatment, following an apparent lag of 2–5 days (see Fig. 3). Mean values for specific QNB binding were 156.5 ± 7.1 fmol per mg protein after 12–15 days of NGF treatment compared with 28.3 ± 6.9 fmol per mg for untreated cells. The increase depended on the presence of NGF and was absent in untreated cultures maintained for the same period at high cell densities (Fig. 3). Interestingly, the major increase in binding, evident by day 5 (Fig. 3), preceded extensive neurite outgrowth.

This study using PC12 cells provides the first evidence for upward regulation of muscarinic ligand binding sites by a chemically defined, developmentally important, neurotrophic substance. These binding sites appear pharmacologically equivalent to muscarinic receptors in normal neural tissues. Their physiological functions both in PC12 cells and in normal tissues, however, require further study. Muscarinic receptors are proposed to influence adrenal cells and sympathetic neurones in several ways⁹⁻¹³, but the most extensive physiological data are derived from the mammalian heart *in vivo*, where presynaptic muscarinic receptors on sympathetic nerve endings augment the effects of vagus stimulation by inhibiting release of catecholamines⁹. PC12 cells thus may provide a useful system to study the biochemical aspects of muscarinic receptor function *in vitro*, as well as the effects of hormones and trophic substances on the expression of these receptors.

We thank Val Slayton for technical assistance. This work was supported by grants from the American Cancer Society (PDT-171) and the National Cancer Institute, DHHS (R23 CA-27808), and the Charles A. King Trust.

Received 18 August 1981; accepted 12 March 1982.

- Greene, L. A. & Tischler, A. S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2424–2428 (1976).
- Greene, L. A. & Rein, G. *Brain Res.* **129**, 247–263 (1977).
- Greene, L. A. & Rein, G. *Brain Res.* **138**, 521–528 (1977).
- Dichter, M. A., Tischler, A. S. & Greene, L. A. *Nature* **268**, 501–504 (1977).
- Ritchie, A. K. *J. Physiol., Lond.* **286**, 541–561 (1979).
- Tischler, A. S. & Greene, L. A. *Lab. Invest.* **39**, 77–89 (1978).
- Greene, L. A. & Rein, G. *Nature* **268**, 349–351 (1977).
- Schubert, D., Heinemann, S. & Kidokoro, Y. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2579–2583 (1977).

- Löffelholz, K. & Muscholl, E. *Naunyn-Schmiedeberg Arch. exp. Path. Pharmac.* **265**, 1–15 (1969).
- Sharma, V. K. & Banerjee, S. P. *Nature* **272**, 276–278 (1978).
- Kayaalp, S. O. & Neff, N. H. *Neuropharmacology* **18**, 909–911 (1979).
- Kayaalp, S. O. & Neff, N. H. *Eur. J. Pharmac.* **57**, 255–257 (1979).
- Adnan, N. A. M. & Hawthorne, J. N. *J. Neurochem.* **36**, 1858–1860 (1981).
- Stallcup, W. B. *J. Physiol., Lond.* **286**, 525–540 (1979).
- Yamamura, H. I. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1725–1729 (1974).
- Kloog, Y., Egozi, Y. & Sokolovsky, M. *Molec. Pharmac.* **15**, 545–558 (1979).
- Yamamura, H. I., Wastick, G., Charg, K. J. & Snyder, S. H. *Proc. West. Pharmac. Soc.* **19**, 13–18 (1976).
- Yamamura, H. I. & Snyder, S. H. *Molec. Pharmac.* **10**, 861–867 (1974).
- Bocchini, V. & Angelletti, P. U. *Proc. natn. Acad. Sci. U.S.A.* **64**, 787–794 (1969).
- Dulbecco, R. & Vogt, M. *J. exp. Med.* **99**, 167–182 (1954).
- Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–672 (1949).
- Patrick, J. & Stallcup, B. *J. biol. Chem.* **252**, 8629–8633 (1977).
- Birdsall, N. J. M. & Hulme, E. C. *J. Neurochem.* **27**, 7–16 (1976).
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. & Randall, R. J. *J. biol. Chem.* **193**, 265–275 (1951).

Adenosine diphosphate induces binding of von Willebrand factor to human platelets

Takayoshi Fujimoto & Jack Hawiger*

Departments of Pathology and Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

Blood platelets seal off leaks in the vessel wall to prevent haemorrhage by a mechanism involving their activation, adhesion to the zone of vessel injury and aggregation¹. This physiologically important haemostatic process requires Factor VIII/von Willebrand factor (vWF), a plasma glycoprotein missing in the genetic, haemorrhagic disorder called von Willebrand disease, in which platelets cannot adhere to the subendothelial layer of the vessel wall². Thus vWF is necessary for anchoring platelets to the vessel wall, and it mediates platelet–platelet interaction. This interaction requires an appropriate receptor on the platelet surface. Until recently, the antibiotic ristocetin was used to induce *in vitro* receptor-mediated binding of vWF to platelets and to cause their aggregation or agglutination³⁻⁷. But although a useful laboratory tool⁸, ristocetin is clearly non-physiological, so we turned our attention to thrombin, which appears in venous blood within 45 s of venipuncture and is a very potent physiological inducer of platelet aggregation. Thrombin causes platelets to change shape from discoid to spiny spheres, and alters the platelet membrane resulting in release of ADP, serotonin and other constituents of platelet storage granules⁹. We recently showed that low concentrations (0.1–0.5 nM) of thrombin, attainable in blood, induce binding of vWF to a specific platelet receptor¹⁰. We now report that the binding of vWF to platelet receptors is also induced by ADP on its addition to platelets. Moreover, ADP released from platelet storage granules by thrombin is, at least in part, responsible for subsequent binding of ^{125}I -labelled vWF.

Platelets from human blood were separated from plasma proteins by stepwise albumin gradient centrifugation and Sepharose 2B gel filtration as described earlier¹¹. All experiments were performed with platelets suspended in HEPES buffer, pH 7.35, containing 0.1% dextrose and 0.35% albumin. Human vWF was purified and labelled with ^{125}I , as before¹⁰. This preparation was rendered essentially free of fibrinogen and fibronectin by passage through an anti-fibrinogen antibody Sepharose 4B affinity column and a gelatin–Sepharose 4B column. Purified vWF gave negative results for fibrinogen content in immunodiffusion and in the staphylococcal clumping test having a detection limit of 0.5 μg fibrinogen per ml ¹². By the use of an even more sensitive radioimmunoassay, we found that 1 mg of purified vWF contained <60 ng of fibrinogen, that is, <0.00006% (w/w).

Binding of ^{125}I -vWF to platelets occurred in 0.5 ml reaction mixture containing 1×10^8 platelets with 5 μM ADP or 0.5 nM

* To whom correspondence should be addressed.

thrombin, to which hirudin and then ^{125}I -labelled vWF had been added. Incubation was at room temperature (22°C) in polypropylene tubes ($12 \times 75\text{ mm}$) without stirring. Incubation was terminated by rapid centrifugation (15,000 r.p.m., 2 min) of platelet suspension through 9:1 (v/v) mixture of dibutylphthalate and Apiezon Oil C (ref. 13) at 60 min, by which time binding had reached equilibrium (see below). The supernatant was sampled for free radioactivity, and bound radioactivity was measured in the platelet pellet which was cut off together with the tip of the tube after rapid freezing of its contents in a mixture of dry ice and acetone. Aggregation of platelets was measured photometrically in a Payton (Buffalo, New York) dual-channel aggregometer at 37°C with constant stirring (900 r.p.m.) according to Born's method¹⁴.

Intact resting platelets, separated from plasma proteins, bound little ^{125}I -vWF. After stimulation with $5\text{ }\mu\text{M}$ ADP, at least three times more ^{125}I -vWF was bound (Fig. 1); this binding approached saturation at $2.8\text{--}5.6\text{ }\mu\text{g}$ ^{125}I -vWF, and equilibrium binding was achieved within 30 min. Whereas approximately 75% of the bound ^{125}I -vWF could be displaced by a 100-fold excess of unlabelled vWF, a similar excess of unlabelled fibrinogen or fibronectin had no effect, suggesting that binding was specific for ^{125}I -vWF.

Induction of binding of ^{125}I -vWF to ADP-stimulated human platelets could be effectively abolished with either the ADP-splitting enzyme, apyrase (5.4 U; see Fig. 1), or an ADP-converting system comprising creatinine phosphate (7.5 mM) and creatine phosphate kinase (3.0 U) (results not shown).

In the above experiments, exogenous ADP was added to platelets before ^{125}I -vWF. However, the dense granules of platelets are a rich source of ADP which is easily releaseable by thrombin⁹. We had earlier shown that thrombin (0.5 nM) induces binding of ^{125}I -vWF to human platelets¹⁰ and so we decided to determine whether there was any relationship between thrombin-induced release of ADP and thrombin-induced binding. The finding that addition of the ADP-splitting

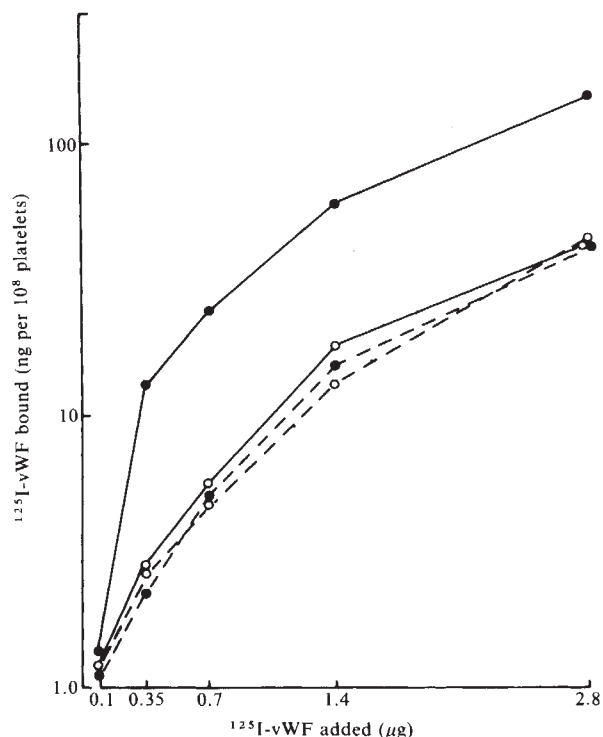


Fig. 1 Binding of ^{125}I -labelled human vWF to platelets treated with $5\text{ }\mu\text{M}$ ADP (●—●); $5\text{ }\mu\text{M}$ ADP plus apyrase (5.4 U; ●—●), buffer (control: ○—○) and buffer plus apyrase (control: ○—○). Platelets were treated with apyrase or buffer for 3 min followed by the addition of ADP. After 3 min ^{125}I -vWF was added. Binding was determined after 60 min incubation at room temperature without stirring.

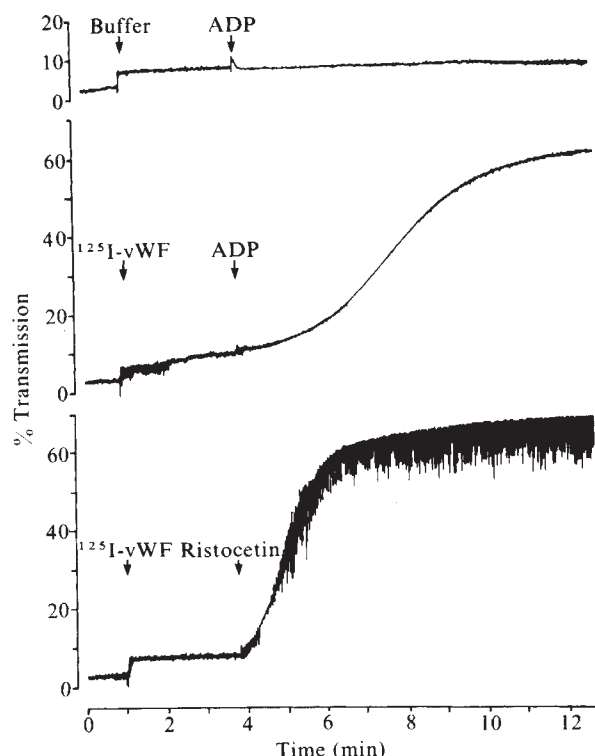


Fig. 2 Aggregation of platelets separated from plasma proteins by ADP and vWF (actual tracings). Upper panel, ADP ($5\text{ }\mu\text{M}$) without vWF; middle panel, ADP + vWF ($30\text{ }\mu\text{g}$); lower panel, ristocetin (1 mg ml^{-1}) + vWF.

enzyme apyrase to thrombin-treated platelets reduced binding by 73% showed that endogenous ADP, released from platelet dense granules and present in the platelet 'environment', was responsible in large part for inducing the binding of ^{125}I -vWF to thrombin-stimulated platelets. ADP-induced binding of ^{125}I -vWF required calcium as both EDTA and EGTA (2 mM) inhibited the interaction.

If binding of multimeric ^{125}I -vWF to the platelet surface is a prerequisite for their aggregation or agglutination, as postulated in the case of ristocetin-induced binding⁷, then platelets stimulated with ADP in the presence of ^{125}I -vWF should also aggregate. Indeed, the induction of vWF binding to human platelets by ADP was paralleled by their aggregation. Human platelets, separated from plasma proteins, did not aggregate in response to either ADP or ristocetin¹¹. However, addition of purified vWF ($30\text{ }\mu\text{g}$) induced aggregation of both ADP- and ristocetin-treated platelets (Fig. 2). As the fibrinogen content of our purified vWF preparation was $<0.00006\%$, and since platelets were not responsive to ADP in the absence of added ^{125}I -vWF, we believe that the observed aggregation of platelets resulted from the induction of the specific receptor for vWF by ADP and the subsequent linking of platelets by bound multimeric vWF.

The previously observed interference by ADP of human vWF-stimulated platelet agglutination in the presence of EDTA and ristocetin¹⁵ can be attributed to changes in the vWF receptor induced by a combination of ADP, ristocetin and EDTA. It is possible that this combination causes a downward shift in the affinity of the receptor for vWF, in contrast to the effect of ADP alone as observed in our system requiring divalent cations.

A major question pertains to nature of the trigger responsible for activation of platelets *in vivo*. Our results support the view of Born¹ that ADP is one such trigger, which by promoting interaction of vWF with platelets allows them to adhere to the vessel wall and to aggregate. *In vivo* activation of platelets by ADP administered intravascularly by microiontophoresis causes them to adhere to the vessel wall¹⁶. Although platelet activation *in vivo* may occur within 10–100 ms (ref. 17), results

obtained in the inherently slower *in vitro* system can still provide a useful new clue to the *in vivo* situation. According to the mechanism we propose here, as vWF is present at the interface between blood and the vessel wall, activation of platelets by ADP results in exposure of receptors for both vWF and fibrinogen¹⁸⁻²³. The ADP-induced receptor for vWF binds this multimeric glycoprotein to form an anchor for the attachment of platelets to the inner zone of the vessel wall and to each other.

Bovine thrombin, hirudin, ADP, apyrase, creatinine phosphate and creatinine phosphate kinase were purchased from Sigma. This work was supported by USPHS grants HL-25935, HL-25107 and HL-27560.

Received 25 September 1981; accepted 1 March 1982.

1. Born, G. V. R. in *Platelets and Thrombosis* (eds Sherry, S. & Scriabine, A.) 113-126 (University Park Press, Baltimore, 1974).
2. Tschopp, T. B., Weiss, H. J. & Baumgartner, H. R. *J. lab. clin. Med.* **83**, 296-300 (1974).
3. Zucker, M. B., Kim, S. J., McPherson, J. & Grant, R. A. *Br. J. Haemat.* **35**, 535-549 (1977).
4. Schneider-Trip, M. D., Jenkins, C. S. P., Kahle, L. H., Sturk, A. & ten Cate, J. W. *Br. J. Haemat.* **43**, 99-112 (1979).
5. Jenkins, C. S. P., Clemetson, K. J. & Luscher, E. F. *J. lab. clin. Med.* **93**, 220-231 (1979).
6. Morisato, D. K. & Gralnick, H. R. *Blood* **55**, 9-15 (1980).
7. Kao, K. J., Pizzo, S. V. & McKee, P. A. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5317-5320 (1979).
8. Howard, M. A. & Firkin, B. G. *Thromb. Diath. haemorrh.* **26**, 362-369 (1971).
9. White, J. G. in *Hematology* (eds Williams, W. J., Bentler, E., Erslev, A. J. & Rundles, R. W.) 1023-1039 (McGraw-Hill, New York, 1980).
10. Fujimoto, T., Ohara, S. & Hawiger, J. *J. clin. Invest.* **69** (in the press).
11. Timmons, S. & Hawiger, J. *Thromb. Res.* **12**, 297-306 (1978).
12. Hawiger, J., Niewiarowski, S., Gurewich, V. & Thomas, D. P. *J. lab. clin. Med.* **75**, 93-108 (1970).
13. Miletich, J. P., Jackson, C. M. & Majerus, P. W. *J. biol. Chem.* **253**, 6908-6916 (1965).
14. Born, G. V. R. *Nature* **194**, 927-929 (1962).
15. Grant, R. A., Zucker, M. B. & McPherson, J. *Am. J. Physiol.* **230**, 1406-1410 (1976).
16. Begent, N. A. & Born, G. V. R. *Nature* **227**, 926-930 (1970).
17. Born, G. V. R. *Ciba Fdn Symp.* **71**, 61-78 (1980).
18. Mustard, J. F., Packham, M. A., Kinlough-Rathbone, R. L., Perry, D. W. & Rogoeczi, E. *Blood* **52**, 453-466, 1978.
19. Marguerie, G. A., Plow, E. F. & Edgington, T. S. *J. biol. Chem.* **254**, 5357-5363 (1979).
20. Bennet, J. S. & Vailre, G. *J. clin. Invest.* **64**, 1393-1401, (1979).
21. Hawiger, J., Parkinson, S. & Timmons, S. *Nature* **283**, 195-197 (1980).
22. Niewiarowski, S., Brudzynski, A. Z., Morinelli, T. A., Brudzynski, T. M. & Stewart, G. *J. J. biol. Chem.* **256**, 917-925 (1981).
23. Peerschke, E. I., Zucker, M. B., Grant, R. A., Egan, J. J. & Johnson, M. M. *Blood* **55**, 841-847 (1980).

Glicentin inhibits gastric acid secretion in the rat

P. Kirkegaard, A. J. Moody, J. J. Holst, F. B. Loud, P. Skov Olsen & J. Christiansen

Department of Surgery D, Glostrup Hospital and Institute of Medical Physiology C, University of Copenhagen, DK-2100 Copenhagen, Denmark and Novo Research Laboratories, DK-2880 Copenhagen, Denmark

Gut glucagon-like immunoreactants (GLIs) are peptides of enteric origin, which have immunoreactivity in common with pancreatic glucagon. First reported in the intestine by Unger *et al.*¹, GLIs are concentrated in the L-cells of the lower small intestine and the colon². The primary structure of gut GLI-1, glicentin, a highly purified form of the major component of gut GLIs³, has recently been established. Glicentin contains 69 amino acid residues, has a molecular weight (M_r) of 8,128 (ref. 4) and contains the entire glucagon sequence^{4,5}, thus establishing it as a member of the hormone family which includes glucagon, secretin, gastric inhibitory polypeptide and vasoactive intestinal polypeptide. Until now there have been insufficient amounts of pure glicentin to test its proposed role in the intestinal phase of the body's response to a meal⁶, but there is evidence which suggests that, like other members of the hormone family to which it belongs, glicentin can inhibit gastric acid secretion⁷⁻⁹. We confirm here that glicentin inhibits pentagastrin-stimulated acid secretion in the rat.

Female Sprague-Dawley rats (~200 g) were used. The rats were fitted with chronic gastric fistulas according to the method of Lane *et al.*¹⁰ and allowed a 2-week post-operative recovery period. The day before the first experiment, the rats were lightly anaesthetized with ether, and a thin polyethylene catheter was placed in a jugular vein for infusion of glicentin, glucagon or

saline. The rats were fasted for 24 h before each experiment, but allowed water *ad libitum*. They were kept in raised mesh-bottom cages to prevent coprophagy.

Ten rats were studied on two separate days. For experimental manipulation, the rats were placed in restraining cages and the stomachs were rinsed out by infusing warm saline through the fistula. After 1 h of spontaneous secretion, pentagastrin (25 μ g per kg; Peptavlon, ICI) was given subcutaneously (s.c.). A bolus injection of either 2 ml saline or 3 μ g per kg glicentin in 2 ml saline was given intravenously (i.v.) followed by infusion of either 0.9% saline (2 ml h^{-1}) or glicentin (600 ng h^{-1}) in 2 ml saline in random order. Another 8 rats received glicentin (120 ng h^{-1}) and 10 rats received glucagon (1,000 ng h^{-1}). All solutions contained 2% human albumin (Behringwerke). The concentration of glucagon-like immunoreactivity in plasma sampled at the end of the infusion were measured, after ethanol-extraction of the plasma (61%), with a radioimmunoassay specific for the glucagon sequence 6-15, which is also present in the glicentin molecule⁵. Gastric juice was collected continuously and sampled at 30-min intervals. Acid output was determined using an autotitrator (Radiometer ABU 13).

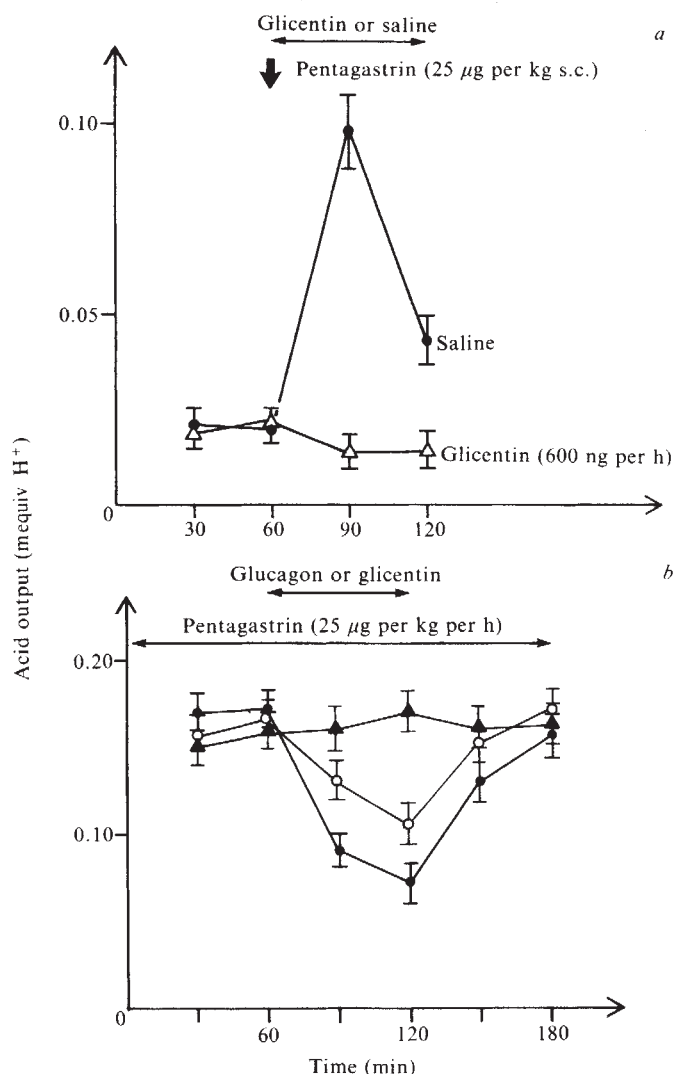


Fig. 1 *a*, Effect of i.v. glicentin infusion on submaximally stimulated gastric acid secretion in the chronic-fistula rat. Acid output (milliequivalents H^+) did not exceed baseline values. Control rats were treated with saline. *b*, Effect of i.v. infusion of glicentin ($\circ$, 120 ng h^{-1} ; $\bullet$, 600 ng h^{-1}) and glucagon ($\blacktriangle$, 1,000 ng h^{-1}) on maximally stimulated gastric acid secretion in the chronic-fistula rat. During infusion of glicentin (both doses), acid output decreased significantly. Glucagon had no effect on acid secretion. For *a* and *b*, each point represents 30 min acid secretion (mean $\pm$ s.e.m. of 10 rats).

In another 10 chronic-fistula rats, pentagastrin (25 µg per kg per h) was infused i.v. for 3 h. After 1 h, a bolus injection of 0.6 or 3 µg per kg glicentin was given and for the next hour, 120 or 600 ng glicentin was infused together with the pentagastrin. For the last hour, pentagastrin alone was infused. The same experiment was performed with glucagon (1,000 ng per h) after a bolus injection of 5 µg per kg. Gastric juice was collected and acid output determined as described above.

The porcine glicentin used was isolated as previously described³ and gave a single band on analytical isoelectrofocusing and polyacrylamide gel electrophoresis. The content of gastric inhibitory polypeptide was <100 p.p.m. (ref. 11) and that of vasoactive intestinal polypeptide was 70 p.p.m. (Moody, A. J. M., unpublished data) as determined by radioimmunoassay.

We observed a marked effect of glicentin on gastric acid secretion. After submaximal stimulation (pentagastrin, 25 µg per kg s.c.) the acid response was completely abolished by the infusion of glicentin (Fig. 1a). After maximal stimulation (continuous i.v. infusion of pentagastrin, 25 µg per kg per h) the acid output decreased in a dose-dependent manner during simultaneous infusion of glicentin. When infusion of glicentin was stopped, there was an immediate increase in gastric acid secretion. Glucagon at 1,000 ng per h had no effect on acid secretion (Fig. 1b). The levels of glucagon-like immunoreactivity in plasma obtained after infusion of glicentin at 600 ng per h and glucagon at 1,000 ng per h, were 541 ± 135 and 428 ± 94 pmol-equivalents l⁻¹, respectively, compared with 72 ± 27 pmol l⁻¹ in rats receiving saline alone (mean $\pm$ s.e.m.).

The present study demonstrates that glicentin is a potent acid secretion inhibitor and the fact that it is effective in low doses supports the view that GLI-1 may act as an enterogastrone in the intestinal phase of gastric acid secretion. As glucagon, given in doses greater than those of glicentin and giving rise to similar concentrations of glucagon-like immunoreactivity, was ineffective, it is unlikely that the action of glicentin is due to that part of its sequence which corresponds to the glucagon molecule.

Received 1 December 1981; accepted 18 March 1982.

1. Unger, R. H., Eisenbraut, A. M., Sims, K., McCall, M. S. & Madison, L. L. *Clin. Res.* **9**, 53-60 (1961).
2. Holst, J. J. *Digestion* **17**, 168-190 (1978).
3. Sundby, F., Jacobsen, H. & Moody, A. J. *Hormone Metab. Res.* **8**, 366-371 (1976).
4. Thim, K. & Moody, A. J. *Regul. Peptides* **2**, 139-151 (1981).
5. Holst, J. J. *Biochem. J.* **187**, 337-343 (1980).
6. Holst, J. J. in *Gut Hormones* (ed. Bloom, S. R.) 383-386 (Churchill-Livingstone, Edinburgh, 1978).
7. Christiansen, J. et al. *Scand. J. Gastroenterol.* **14**, 161-166 (1979).
8. D Sa Barro, A. A. B. & Buchanan, K. D. *Gut* **18**, 877-881 (1977).
9. Coudray, A. M. & Rosselin, G. *Biosci. Rep.* **1**, 151-155 (1981).
10. Lane, A., Ivy, A. C. & Ivy, E. M. *J. Physiol.* **192**, 221-228 (1957).
11. Lauritsen, K. B. & Moody, A. J. *Diabetologia* **14**, 149-153 (1978).

A sequence of mitochondrial DNA is associated with the onset of senescence in a fungus

Corinne Vierny, Anne-Marie Keller, Odile Begel & Léon Belcour

Centre de Génétique Moléculaire, Central National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France

In *Podospira anserina*, a strict aerobic fungus, senescence¹ (arrest of vegetative growth of the mycelium) is correlated with severe modifications of the mitochondrial DNA similar to those observed in the *rho*⁻ mutants of yeast². Senescent cultures yield a specific DNA (SEN-DNA) consisting of small circular molecules arranged in multimeric series of sizes^{3,4} and which results from the amplification of either one of two non-overlapping segments of the mitochondrial DNA^{5,6}. One region 0.89 µm long [2.6 kilobases (kb)] is most frequently amplified, in different strains of *P. anserina* and also in subcultures of a

strain reaching senescence independently⁶. In 'young' cultures of the mating type-minus wild-type strain *s mt*⁻, the 2.6 kb region exists both as an integral part of the mitochondrial chromosome and as rare free sequences⁶. The results presented here show that (1) the free form is absent from two strains, *s mt*⁺ and *s mt*^{- cap}¹, which both display increased longevity and which differ respectively from the *s mt*⁻ strain by a nuclear allele at the *mt* locus and by a mitochondrial mutation conferring resistance to chloramphenicol; and (2) in the mutant *mex*¹, which develops beyond the edge of arrest of a senescent mycelium, that is, escapes senescence, the mitochondrial DNA precisely lacks the 2.6 kb region. The results suggest that excision and/or amplification of the 2.6 kb sequence is the cause of senescence.

We previously reported that senescent cultures of *P. anserina* contain one of two types of SEN-DNA differing by their density⁶. The first type has a density of 1.694 g cm⁻³ (identical to that of the normal 'young' mitochondrial DNA) and exhibits a variable monomer unit comprising 4-6.3 kb. Cross-hybridization revealed that all SEN-DNAs of density 1.694 g cm⁻³ so far obtained share a sequence of mitochondrial DNA of ~2.3 kb and thus belong to the same family. The second type of SEN-DNA has a density of 1.699 g cm⁻³, a size of 2.6 kb and a sequence that has been reproducibly recovered from independent senescent cultures grown in this and other laboratories^{4,6,7}. Hybridization of SEN-DNAs with the mitochondrial DNA of a young culture of the *s mt*⁻ strain digested by *Hae*III (see Table 1 for a description of the strain) shows that both types of SEN-DNAs result from the amplification of particular regions of the mitochondrial chromosome^{5,6}. However, whereas the 1.694 g cm⁻³ SEN-DNA hybridizes only with fragments of the 'young' mitochondrial DNA visible on the gel, the 1.699 g cm⁻³ SEN-DNA hybridizes to a 2.6 kb position as well, at which no visible fragment is found (see Fig. 1a and its legend). We have therefore assumed that the 2.6 kb sequence is present in young *s mt*⁻ cultures not only as an integral part of the mitochondrial chromosome but also as rare free molecules⁶. Three properties of the 1.699 g cm⁻³ SEN-DNA—its specific density, its peculiar pattern of hybridization with the 'young' *s mt*⁻ mitochondrial DNA and its repeated appearance in senescent cultures of independent origin—distinguish it from 1.694 g cm⁻³ SEN-DNAs and suggest that it has episome-like properties. In any case, whatever the basis for the difference between the two types of sequences, it is clear that senescence is always correlated with the amplification of one of these types of sequence. This correlation raises the question of the cause-effect relationship between senescence and the presence of the amplified sequences, and it is tempting to assume that the circular amplified molecules constituting the SEN-DNAs correspond to the infectious cytoplasmic factor responsible for senescence and identified by Marcou⁸.

Two sets of new data on the 2.6 kb sequence provide good support for the idea that senescence actually results, at least in some cases, from the amplification of this sequence. (1) The *Hae*III-digested mitochondrial DNAs extracted from young cultures of the two strains *s mt*⁺ and *s mt*^{- cap}¹ (whose life spans are longer than that of strain *s mt*⁻; Table 1) were hybridized with the 1.699 g cm⁻³ SEN-DNA. In both cases, no

Table 1 Life spans of the different strains used

Genotype		Life span (cm)
Nuclear	Mitochondrial	
<i>s mt</i> ⁻	<i>cap</i> ^s	14
<i>s mt</i> ⁺	<i>cap</i> ^s	40
<i>s mt</i> ^{- cap} ¹	<i>cap¹</i>	100
<i>s mt</i> ⁻	<i>cap</i> ^{s mex} ¹	>110

* Strain *s mt*⁺ is isogenic with strain *s mt*⁻ except for the mating-type locus and the two strains are cytoplasmically identical.

† Strain *s mt*^{- cap}¹ is strictly isogenic with strain *s mt*⁻ from which it derives, except for the mitochondrial mutation *cap*¹.

labelled material was recovered at the 2.6 kb position (Fig. 1A, parts *b* and *c*). This indicates that in the strains *s mt*⁺ and *s mt*⁻ *cap*⁺1, the 2.6 kb sequence does not exist in a free form but is present only as an integrated sequence. (2) A mutant *mex1*, having a modified lifespan, was obtained and studied. This mutant was selected as a sector of growing mycelium that developed beyond the arrested edge of a *s mt*⁻ senescent culture. Such revertants of the senescent state are uncommon. The mutant *mex1* exhibits a slow growth rate but since its isolation it has been growing steadily and has shown no symptoms of senescence. Preliminary genetic analysis of mutant *mex1* indicated that it segregates in a non-mendelian way, which suggested that the mutation might be located in the mitochondrial DNA. Restriction analysis of the mitochondrial DNA of this mutant confirmed this hypothesis. Figure 2 shows the restriction pattern (after digestion by *Bgl*II) of the mitochondrial DNA of mutant *mex1* compared with that of the wild-type strain

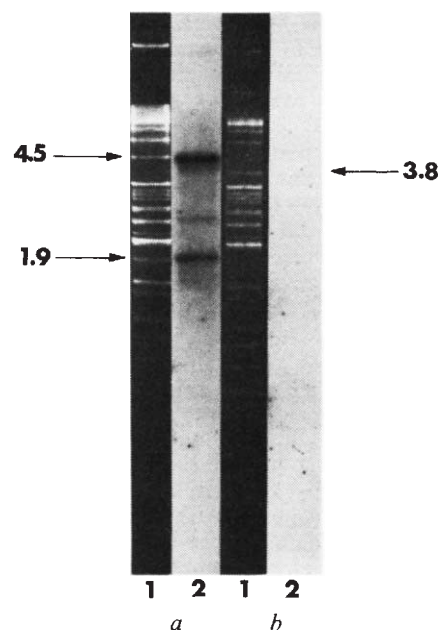


Fig. 2 Lanes 1, electrophoretic patterns of mitochondrial DNA of strains *s mt*⁺ (*a*) and *s mt*⁻ *mex1* (*b*) digested by *Bgl*II. Lanes 2, autoradiographs after transfer of the fragments to nitrocellulose filter and hybridization with a ³²P-labelled 1.699 g cm⁻³ SEN-DNA obtained by nick-translation (*a*, strain *s mt*⁺; *b*, strain *s mt*⁻ *mex1*). Note the monomer unit of the 2.6 kb SEN-DNA possesses one site for the restriction enzyme *Bgl*II. Sizes in kb.

from which it derives. The *mex1* mitochondrial DNA lacks two bands of respectively 4.5 and 1.9 kb and presents an additional band of 3.8 kb, thus *mex1* has a deletion of ~2.6 kb. Interestingly, the missing fragments are precisely those previously shown to hybridize with the 1.699 g cm⁻³ SEN-DNA. The number and the nature of the fragments missing in *mex1* and the estimated size of the deleted fragment obviously suggested that the mutation *mex1* corresponded to the deletion of the 2.6 kb sequence which is often amplified in senescent cultures. This was confirmed by hybridization experiments on the mitochondrial DNA of both young wild-type and *mex1* cultures after digestion by *Bgl*II, using a 1,699 g cm⁻³ nick-translated SEN-DNA as ³²P-labelled radioactive probe. The results of this experiment (Fig. 2) show that the mitochondrial DNA of mutant *mex1* does not exhibit any homology with the 1.699 g cm⁻³ SEN-DNA. Similar hybridization experiments in which the mitochondrial DNA *mex1* is digested by *Hae*III (Fig. 1A, *d*) or *Bam*HI yield the same result. It can be concluded that in the mutant *mex1*, a sequence containing at least the monomer of the 1.699 g cm⁻³ SEN-DNA has been deleted. The electrophoretic pattern of the *Hae*III- (or *Bgl*II-) digested mitochondrial DNA of mutant *mex1* indicated the absence of any modifications in the region capable of giving rise to the 1.694 g cm⁻³ family of SEN-DNAs. This was confirmed by experiments in which *Hae*III-digested *mex1* and *s mt*⁻ mitochondrial DNAs were hybridized with 1.694 g cm⁻³ ³²P-labelled nick-translated SEN-DNA (monomer unit = 6.3 kb); the same hybridization pattern was obtained in both cases (Fig. 1B).

The data presented above show clearly that the 2.6 kb sequence has a central, although perhaps not unique, role in the determination of senescence. The existence of a mutant missing this sequence, which after more than 1 m of growth (see mean life span of the corresponding wild-type strain in Table 1) had not yet become senescent, suggests that excision and/or amplification of the 2.6 kb sequence are the cause of senescence or at least necessary for its onset. This hypothesis is supported by hybridization experiments which reveal that the state in which this sequence exists in young cultures (free and integrated in strain *s mt*⁻, integrated only in strains *s mt*⁺

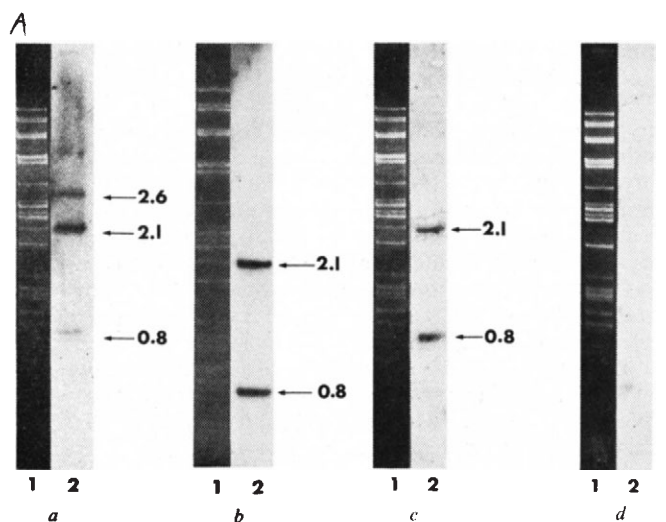


Fig. 1 A, Southern hybridizations between *Hae*III-digested 'young' mitochondrial DNAs of different strains and ³²P-labelled 1.699 g cm⁻³ nick-translated SEN-DNA. The conditions for preparation and endonuclease digestion of DNA, gel electrophoresis and transfer of DNA to nitrocellulose filters were as previously described⁵. The labelling of DNA by nick-translation was performed as described in ref. 12. For each DNA studied, the electrophoretic pattern and the autoradiograph of the filter after hybridization are shown in lanes 1 and 2 respectively. Sizes in kb. *a*, Strain *s mt*⁺; *b*, strain *s mt*⁻; *c*, strain *s mt*⁻ *cap*⁺1; *d*, strain *s mt*⁻ *mex1*. Note that the molecules which constitute the 2.6 kb SEN-DNA are circular and that the monomer unit of this SEN-DNA has a single *Hae*III site. Only one fragment of 2.6 kb is thus recovered after digestion of this SEN-DNA by *Hae*III. B, Southern hybridization between *Hae*III-digested 'young' mitochondrial DNAs of strains *s mt*⁺ (*a*) and *mex1* (*b*) and ³²P-labelled 1.694 g cm⁻³ nick-translated SEN-DNA (monomer unit = 6.3 kb). In each case, the electrophoretic pattern and the autoradiograph of the filter after hybridization are shown in lanes 1 and 2 respectively. Sizes in kb. Note that the monomer unit of the 6.3 kb SEN-DNA possesses two *Hae*III sites.

and *smt⁻cap¹* and absent in strain *mex1*) seems correlated with the longevity of the strains; the higher the probability of appearance of the free 2.6 kb sequence, the shorter the longevity of the strain. The fact that the strains *smt⁺* and *smt⁻cap¹* are isogenic except for a nuclear (*mt⁺/mt⁻*) or a mitochondrial (*cap¹/cap¹*) gene implies that these might act on, for example, the probability of excision or the efficiency of amplification of the 2.6 kb sequence. Nothing is yet known about the mechanisms responsible for the appearance of the 1.699 g cm⁻³ SEN-DNA. However, the presence of a population of molecules lacking the 2.6 kb sequence in the residual mitochondrial DNA of senescent cultures displaying this amplified sequence⁹ (our results, in preparation) suggests that excision rather than replication is the primary event necessary for further amplification of this sequence. The existence of a mutant lacking the 2.6 kb sequence indicates that the sequence does not code for a product essential for cell survival. The reduced growth rate of mutant *mex1* suggests, however, that the sequence is physiologically important, perhaps, as already

postulated⁶, in initiating the replication of the mitochondrial chromosome. As more than one replication origin is likely, as demonstrated in yeast^{10,11}, the mitochondrial chromosome would still be able to replicate in the absence of the 2.6 kb sequence but at a slower rate.

The phenomenon of amplification during senescence is not exclusive to the 2.6 kb sequence. Another family of sequences sharing a common region can also be amplified and, depending on the senescent culture, this second type of sequence is amplified either alone^{5,6} or in association with the 2.6 kb sequence (our unpublished results). In the latter case, it is not yet clear whether the amplification of the two types of sequence is related, although preliminary results of other senescence-resistant mutants suggest that the absence or rearrangement of the second family of sequences also prevents the appearance of senescence.

This work was supported in part by INSERM grants 794 0062 and 0038279114. We thank Dr Janine Beisson for her careful and critical reading of the manuscript.

Received 19 January; accepted 17 March 1982.

- Rizet, G. *C. r. heb. Séanc. Acad. Sci., Paris* **237**, 838–840 (1953).
- Slonimski, P. P. & Lazowska, J. *Mitochondria* **77** (ed. Bandlow, W.) 39 (De Gruyter, Berlin, 1977).
- Stahl, U., Lemke, P. A., Tudzynski, P., Kück, U. & Esser, K. *Molec. gen. Genet.* **162**, 341–343 (1978).
- Cummings, D. J., Belcour, L. & Grandchamp, C. *Molec. gen. Genet.* **171**, 239–250 (1979).
- Jamet-Vierny, C., Begel, O. & Belcour, L. *Cell* **21**, 189–194 (1980).

- Belcour, L., Begel, O., Mossé, M. O. & Vierny, C. *Curr. Genet.* **3**, 13–21 (1981).
- Stahl, U., Kück, U., Tudzynski, P. & Esser, K. *Molec. gen. Genet.* **178**, 639–646 (1980).
- Marcou, D. *Annls Sci. nat. bot.* **11**, 653–764 (1961).
- Kück, U., Stahl, U. & Esser, K. *Curr. Genet.* **3**, 151–156 (1981).
- Blanc, H. & Dujon, B. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3942–3946 (1980).
- Gourshot, R., De Zamaroczy, M., Baldacci, G. & Bernardi, G. *Curr. Genet.* **1**, 173–176 (1980).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).

Structural analysis of insertion sequence IS5

Manfred Kröger & Gerd Hobom

Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D-7800 Freiburg i.Br., FRG

Among the transposable DNA elements that have been detected in bacterial genomes and plasmids^{1–3}, the small insertion sequence (IS) elements have been recognized as a class of autonomous, basic units of transposition. IS elements are not only fully competent of transposition and able to cause several other related chromosomal aberrations such as deletion of adjacent DNA segments^{4,5} and integration of circular plasmid DNAs⁶, but have also proved to be the active units of transposition in the larger composite transposon (Tn) elements^{1–3,7–9}. In addition, IS and Tn elements interfere with local transcription patterns in the host genome by causing polarity^{10,11}. Structural and functional analysis of the IS elements should therefore elucidate both the basis for the autonomy of a genetic unit as small as 1,000 base pairs (bp), and the transposition mechanism, including its inherent step of DNA duplication¹². We have now determined the 1,195 bp sequence of the transposable DNA element IS5 in phage λ KH100. Near-terminal signal structures appear to protect the central region from outside-in transcriptions and translations, and render IS polarity as a direct consequence of IS autonomy. The nucleotide structure shows an unusual, highly compact organization of two completely overlapping, antiparallel genes and correlated control sequences.

IS5 was first identified among clear plaque mutants of bacteriophage λ (λ KH100), and was accordingly localized to the C-terminus of the *cI* repressor gene^{13,14}. IS5 was estimated to be 1,200–1,250 bp long, and electron microscopy revealed short inverted sequence repetitions at its ends. The IS5 unit of λ KH100 seemed different from IS1–IS4, but identical with two other insertions respectively in the bacterial *argB* gene of specialized transducing phage λ 13dargB2 and in phage Mu DNA¹⁵. Both ends of IS5 are known to be deletogenic¹⁵ and to stimulate general recombination within adjacent DNA

segments¹⁶. IS5 also appears to exert polar effects in at least one orientation^{17,18}.

We have used λ KH100 DNA as a primary source of IS5, and have constructed a hybrid plasmid pHL118 by cloning a 1,720 bp *Hind*III fragment into a pBR313-derived vector plasmid (see ref. 12). The λ KH100 *Hind*III fragment contains all of the IS5 DNA as well as 564 bp of λ DNA (λ : 37,727–38,291), and it was used for DNA sequencing according to the Maxam and Gilbert method¹⁹ (see Fig. 1). Mapping of the resulting restriction fragments has provided independent proof of the existence and location of a large number of endonucleolytic recognition sites (see Fig. 2). Both junctions and the λ KH100 DNA sequences adjacent to IS5 have also been sequenced and found to be consistent with published data^{20,21} obtained from other λ phage or plasmid DNAs. IS5 is only slightly smaller than was estimated by electron microscopy¹⁵, and the position of IS5 in the λ KH100 sequence (codon 207 of the 232 codon *cI* gene, see Fig. 2a) is only slightly to the left of two genetic determinations^{13,16}.

When IS5 is inserted into λ cI DNA, 4 bp of the wild-type sequence (target site TTAG; λ :38,150–38,153) are directly duplicated to the left and right of both IS5 ends, and are connected to the two IS5 terminal inverted repeat sequences of 16 bp (with a single mismatch at position 13) (see Fig. 2a). No internal repetition of the inverted repeat sequence has been detected. It is now clear that a partial sequence of phage Mu445–8 DNA²² includes an identical 99 bp fragment from the right *Eco*RI end of IS5 (with a single deviation at IS5:1,171), followed by an identical target site sequence in opposite orientation (CTAA). A 4 bp target site duplication has not previously been observed for the integration of prokaryotic transposable elements, but has been reported in the integration of Moloney strains of murine leukaemia and sarcoma viruses in eukaryotic DNA^{23,24}.

Most of the internal sequence of IS5 is dominated by a very long open reading frame, which extends in right to left direction from position IS5:1,127 to IS5:147 and codes for a 326 amino acid, predominantly basic protein of molecular weight 37,831^{±60}. The IS5 large gene coding region is preceded by a Shine–Dalgarno²⁵ initiation sequence (the most likely fit is shown in Fig. 2f). A computer analysis also identified potential promoter and terminator signals²⁶ before (IS5:1,175–1,139)

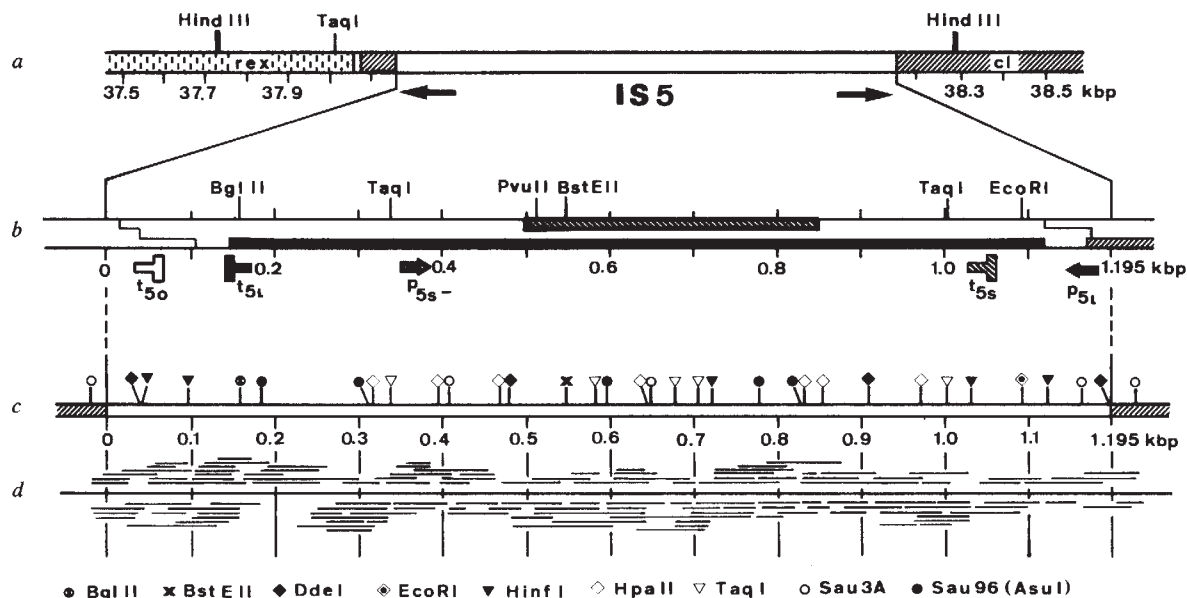


Fig. 1 Restriction map of the *rex-cl* region of λ KH100 DNA and sequencing chart for insertion element IS5 DNA. *a*, Position of insertion IS5 in the *cl* gene of λ KH100. Hatched regions represent IS5-adjacent λ DNA; sequence positions in λ DNA refer to the *EcoRI* restriction site located at the origin of replication (position 40,000). The marked *HindIII* fragment was cloned into a pBR313-derived vector plasmid and propagated as described previously³³. *b*, Genetic map of insertion element IS5 from the complete nucleotide sequence in Fig. 2 and from a functional analysis of its terminator and promoter signals of transcription (see Fig. 3 and text). The black bar indicates the coding region for the large IS5 protein, the hatched bar corresponds to the IS5 small protein reading frame. Outside-in translation according to the diverted λ cl reading frame is indicated as a lightly hatched bar at the right IS5 boundary, and the near-terminal stop codons for the other five reading frames are indicated at the left and right termini respectively. Three transcriptional terminator signals (t_{50} , t_{51} , t_{5s}) and two promoter sequences (p_{51} , p_{5s}) are indicated in their tentative locations. The restriction sites used during functional studies (see Fig. 3 legend) are shown above the genetic map. *c*, Restriction map of IS5 DNA, indicating the restriction sites used in generating 5'-labelled fragments for DNA sequencing: *BglII*, *BstEII*, *DdeI*, *EcoRI*, *HinfI*, *HpaII*, *Sau3A*, *Sau96* and *TaqI*. A key for restriction site symbols is given at the bottom of the figure. *d*, Summary of DNA sequence results obtained for IS5 DNA. Every stretch of independently determined DNA sequence is indicated by a single horizontal bar covering the corresponding region of unequivocal reading; segments of *l*-strand DNA are given above, *r*-strand segments below the central line. Note that almost the entire element has been sequenced using both strands, and large sections have been sequenced repeatedly, from different endonuclease cuts, as indicated by piles of horizontal bars. DNA sequencing was by the Maxam-Gilbert method, using the A > C reaction to determine adenine positions¹⁹. All isolated fragments were purified by butanol extraction³⁴ before both the kinase reaction and chemical degradation steps.

and after (IS5:166–143) the large gene coding sequence respectively.

Although none of the other two reading frames in right to left direction is open for any considerable length, the IS5 DNA sequence appears to have the coding potential for a second protein in left to right direction, from position IS5:525 to IS5:851. The second open reading frame, which is in register with the sequence coding for the larger IS5 protein, is predicted to code for a 108 amino acid, predominantly apolar protein of molecular weight 11,069⁺¹¹₋₁₀. Analogous to the IS5 large gene, the IS5 small gene also is preceded by a ribosome recognition sequence (see Fig. 2g; an alternative AUG start codon in the same reading frame, at IS5:495, is not affiliated with a recognizable Shine-Dalgarno sequence), and, according to computer search, has correlated promoter and terminator signals (at IS5:355–401 and 1,019–1,054) before and after the coding sequence.

To measure IS5 promoter and terminator activities *in vivo*, we and Rak *et al.*²⁷ have individually subcloned into hybrid plasmids certain fragments containing IS5 signal elements. Our analysis of the IS5 terminator fragments (Fig. 3) confirmed computer-predicted locations and orientations, within the present limits of minimization. In addition to the large gene and small gene terminators (t_{51} and t_{5s}), a termination signal has been detected near the left terminus of IS5 in outside-in orientation (t_{50} ; IS5:31–71) which is consistent with the observed polarity of the element¹⁸. No analogous terminator has been found near its right end, however. The position of the IS5 large gene promoter (p_{51}) has been confirmed by Rak *et al.*, while the IS5 small gene promoter (p_{5s}) was located at IS5: 460–495 in their cloning analysis. Expression of IS5 large and small proteins has also been detected in minicells²⁷.

No other example of the IS5 type of DNA dual coding capacity has been reported. Computer screening of published IS DNA sequences—IS2 (ref. 28), IS4 (ref. 29) and IS903 (ref.

30) has, however, revealed quite similar reading frame organizations. In addition to a dominating large reading frame of more than 300 codons, there is always a second reading frame of 108–143 codons in the opposite orientation. Note that the large and small reading frames are consistently found in register, and that the (presumptive) small gene is always completely covered by the large gene. The sequence of IS1^{31,32} shows a similar pattern, although both the large and small reading frames are short: 167 and 89 codons respectively. Similar to IS5 (Fig. 2g), the small gene Shine-Dalgarno recognition sequences always appear as part of mRNA hairpin structure covering the smaller reading frame initiation region, and may be indicative of a generally reduced expression. There is no evidence, however, in any other element of promoter or terminator transcription signals or of dual protein expression. The role of such a compact antiparallel arrangement in genetic regulation or other functions is unknown.

An IS element preserves its autonomy by a genetic organization which avoids regularity dominance by adjacent genes or transcription units on the host DNA. There are therefore near-terminal stop codons in all six reading frames (marked in Fig. 2b) to avoid formation of fusion proteins under foreign control. In addition, near terminal termination signals of transcription, and therefore polarity, may be a direct consequence of IS autonomy. Whereas the structure of IS4, which has two outside-in termination signals²⁹, fully agrees with this prediction, IS5 does not carry a termination signal at its functionally rather crowded right-hand end. IS autonomy may be protected in other ways, however (compare Fig. 2f). The biological significance of the different organization of the two ends of IS5 is not known.

We thank B. Rak for communicating and discussing results before publication, and B. Traub and G. Schlingmann for technical assistance. λ KH100 was kindly provided by W. Szybalski.

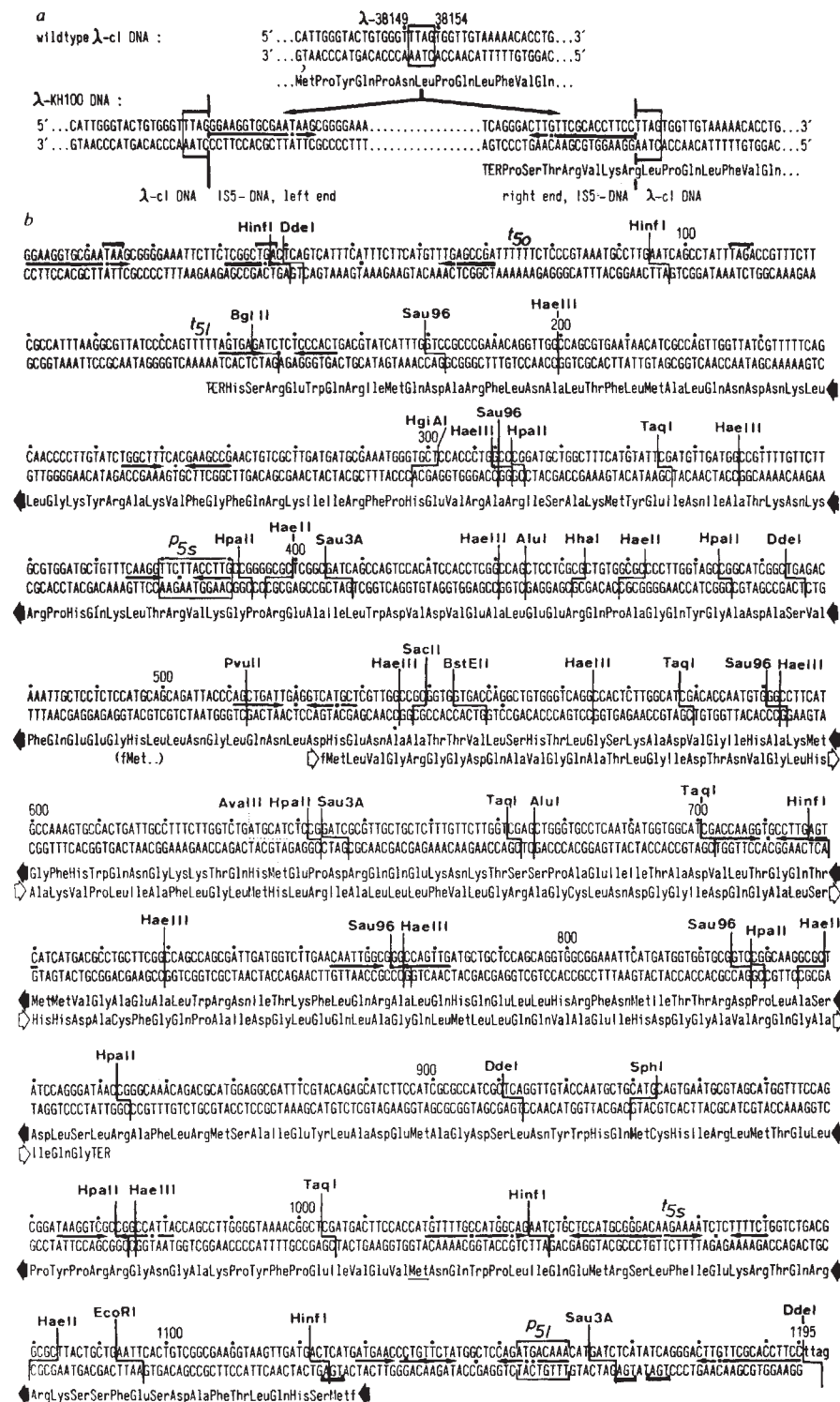
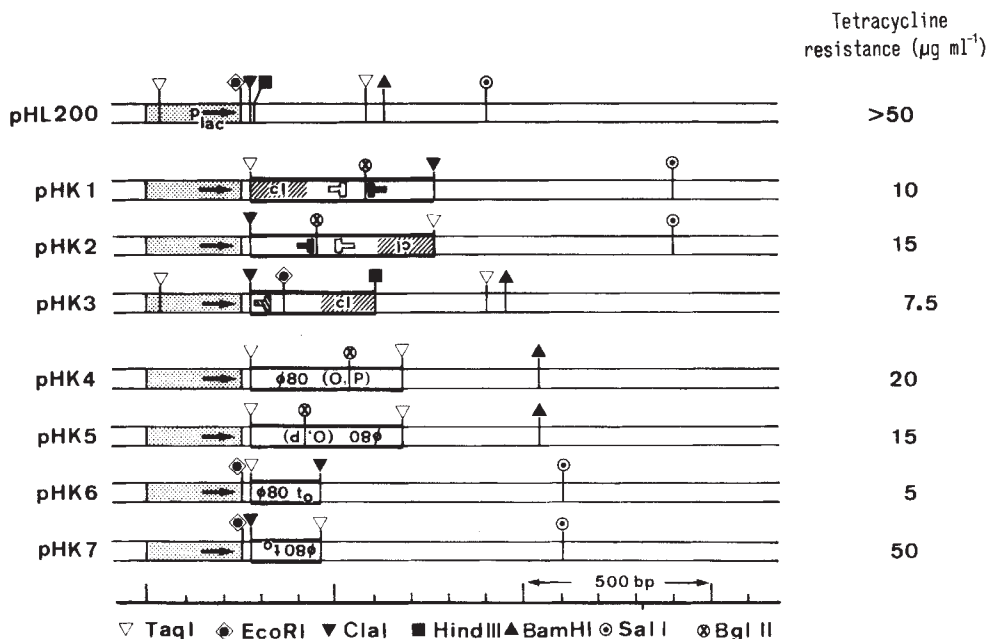


Fig. 2 DNA sequence of insertion element IS5 in λ KH100 DNA. **a**, Identification of the IS5 integration site in the coding sequence of the cI repressor gene of phage λ KH100. The 4 bp target site duplication is boxed, and the 16 bp terminal inverted repeat sequences (with a single mismatch) indicated by arrows. A proportion of the λ wild-type and of λ KH100 cI protein sequences are shown below the DNA sequences. According to the genetic map of phage λ , the termini of IS5 have been named left and right end respectively. **b**, Complete nucleotide sequence of IS5. Numbering starts from the left end. Amino acid sequences predicted from the DNA sequences of the large (right to left; closed arrows) and small (left to right; open arrows) genes of IS5 are given in separate lines below the DNA sequence. Two potential alternative initiation sites for the large and small genes are indicated at positions 1,019 and 495 respectively. Cleavage sites for a collection of restriction enzymes that have been used for fragment mapping or DNA sequencing are marked at the respective nucleotide position above the DNA sequence. (In addition to experimentally proved restriction cuts, the recognition sequences for *Ava*III, *Hae*II and *Sph*I are also indicated.) Arrows between the DNA strands designate inverted sequence repetitions. Brackets refer to near-terminal stop codons with only the one closest to the terminus indicated for each of the six reading frames (excluding those in the target site duplication). Both rightward and leftward promoters are indicated by boxes at their tentative locations. Methylated cytidine *Eco*RII/*Bst*NI recognition sequences are indicated by squares; cleavage at the *Sau*96/*Hae*III restriction site at IS5:310 occurred in low yields, apparently due to an overlap with a methylated C residue. **c-e**, DNA secondary structures predicted for transcription terminator signals of IS5. Terminators for rightward outside-in transcription, for the IS5 large gene (leftwards) and the IS5 small gene transcription (rightwards), respectively, are drawn in a non-coding strand hairpin/coding strand 'tail' sequence manner (see refs 33, 35). Numbers refer to the respective positions in the IS5 DNA sequence. **f**, g, RNA secondary structures predicted to occur at the initiation regions for the IS5 large and small gene proteins respectively. The RNA hairpin structure in *f* can only be formed in a (long) outside-in transcription messenger and not within a (short) mRNA started at *p*₅₁. According to a best fit with a bacterial promoter consensus sequence²⁶, the 5' end of the *p*₅₁ mRNA is located at the position marked by an arrow (IS5:1,142), which would preclude formation of an initiation region hairpin structure. AUG initiator codons are boxed; bars along the sequences refer to Shine-Dalgarno recognition sequences²⁵.

Fig. 3 Hybrid plasmid functional analysis of IS5 terminator fragments. For all cloning experiments the same vector DNA, pHL200, has been used. This vector is a derivative of pBR322 and carries to the left of the *EcoRI* site a *lacUV5* promoter fragment ~286 bp long (stippled area with arrow pointing in direction of RNA synthesis). The starting material to construct pHL200, pEx(lac)UV5-150, was kindly provided by H. Weiher, B. Zink and H. Schaller. Insertion of foreign DNA fragments between the *lac* promoter and downstream tetracycline-resistance genes made use of the *Clal* and *HindIII* cloning sites. The tetracycline promoter of pBR322 extends across these restriction sites, but is destroyed by insertion. Care had to be taken, however, not to reconstitute the tetracycline promoter sequence through insertion of sequences similar to the missing -35 region. The levels of tetracycline resistance have been measured on the same freshly prepared agar plates for the whole set of clones together with a pBR322-containing strain (>50 $\mu\text{g ml}^{-1}$ tetracycline) and an 'empty' strain of *Escherichia coli* K-12-490A (3 $\mu\text{g ml}^{-1}$ tetracycline) as controls.

The respective limiting concentrations of tetracycline are given on the right. DNA fragments incorporated into pHL200 are shown in heavier lines. The cleavage sites of restriction enzymes used for cloning or plasmid analysis are indicated by symbols defined at the bottom of the figure. The λ KH100 DNA IS5 left terminal *TaqI* fragment (λ :38,000-IS5:335) has been cloned into the *Clal* site of pHL200 in both orientations which correspond to the outside-in (t_{5a}) and large gene terminator (t_{5b}) respectively. Two sets of clones with different tetracycline resistance levels have been obtained: pHK1 and pHK2. A λ KH100 DNA IS5 right terminus containing *TaqI*-*HindIII* fragment (IS5:1,005- λ :38,291) corresponding to the small gene terminator, t_{5a} , has been cloned into the *Clal*-*HindIII* gap of pHL200. The resulting hybrid plasmid pHK3 showed a strong reduction of the tetracycline resistance level. The control clones have been constructed in a similar way using two ϕ 80 DNA *TaqI* fragments for insertion into the *Clal* site of pHL200: ϕ 80:40,107-40,514 which contains no transcription signal, but two long ϕ 80-*O*, *P* reading frames either in sense or in antisense orientation (pHK4, pHK5), which extend into and probably interfere with the tetracycline resistance initiation region, and ϕ 80:39,406-39,584 containing the t_0 signal in direct or inverse orientation (pHK6, pHK7). The position of the IS5 transcriptional terminator signals and the adjacent λ sequences are indicated according to Fig. 1.



Note added in proof: Since submission of this paper two independent sequence determinations of IS5 DNA from different isolates have been published. They gave an identical³⁶ or nearly identical³⁷ (two deviations) result.

Received 16 March 1981; accepted 18 March 1982.

- Starlinger, P. *Plasmid* 3, 241-259 (1980).
- Calos, M. P. & Miller, J. H. *Cell* 20, 579-595 (1980).
- Kopecko, D. J. *Progr. molec. subcell. Biol.* 7, 135-234 (1980).
- Reif, H. J. & Saedler, H. *Molec. gen. Genet.* 137, 17-28 (1975).
- Kleckner, N., Reichardt, K. & Botstein, D. *J. molec. Biol.* 127, 89-115 (1979).
- Davidson, N., Deonier, R. C., Hu, S. & Ohtsubo, E. in *Microbiology* 1974 (ed. Schlessinger, D.) 56-68 (American Society for Microbiology, Washington DC, 1975).
- MacHattie, L. A. & Jackowski, J. B. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 219-228 (Cold Spring Harbor Laboratory, New York, 1977).
- Kleckner, N. & Ross, D. G. *Cold Spring Harb. Symp. quant. Biol.* 43, 1233-1246 (1979).
- Iida, S., Meyer, J. & Arber, W. *Cold Spring Harb. Symp. quant. Biol.* 45, 27-43 (1981).
- Jordan, E., Saedler, H. & Starlinger, P. *Molec. gen. Genet.* 132, 353-363 (1968).
- Adhya, S. L. & Gottesmann, M. E. *A. Rev. Biochem.* 47, 967-996 (1978).
- Lusky, M., Kröger, M. & Hobom, G. *Cold Spring Harb. Symp. quant. Biol.* 45, 173-176 (1981).
- Blattner, F. R. *et al. Virology* 62, 458-471 (1974).
- Szybalski, W. in *DNA Insertion Elements, Plasmids, and Episomes* (eds Bukhari, A. I., Shapiro, J. A. & Adhya, S. L.) 583-590 (Cold Spring Harbor Laboratory, New York, 1977).
- Chow, L. T. & Broker, T. R. *J. Bact.* 133, 1427-1436 (1978).
- Lieb, M. *Gene* 12, 277-280 (1980).
- Glansdorff, N. & Sand, G. *Biochim. biophys. Acta* 108, 308-311 (1965).
- Crabeel, M., Charlier, D., Cunin, R. & Glansdorff, N. *Gene* 5, 207-232 (1979).
- Maxam, A. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* 74, 560-564 (1977).
- Sauer, R. T. *Nature* 276, 301-302 (1978).
- Landmann, J., Kröger, M. & Hobom, G. *Gene* (in the press).
- Kahmann, R. & Kamp, D. *Nature* 280, 247-250 (1979).
- Dhar, R., McClements, W. L., Enquist, L. W. & van de Woude, G. F. *Proc. natn. Acad. Sci. U.S.A.* 77, 3937-3941 (1980).
- Shoemaker, C. *et al. Proc. natn. Acad. Sci. U.S.A.* 77, 3982-3996 (1980).
- Shine, J. & Dalgarno, L. *Proc. natn. Acad. Sci. U.S.A.* 71, 1342-1346 (1974).
- Rosenberg, M. & Court, D. A. *Rev. Genet.* 13, 319-353 (1979).
- Rak, B., Lusky, M. & Hable, M. *Nature* 297, 124-128 (1982).
- Ghosal, D., Sommer, H. & Saedler, H. *Nucleic Acids Res.* 6, 1111-1122 (1979).
- Klaer, R. *et al. Cold Spring Harb. Symp. quant. Biol.* 45, 215-224 (1981).
- Grindley, N. & Joyce, C. M. *Cold Spring Harb. Symp. quant. Biol.* 45, 125-133 (1981).
- Ohtsubo, H. & Ohtsubo, E. *Proc. natn. Acad. Sci. U.S.A.* 75, 615-619 (1978).
- Johnsrud, L. *Molec. gen. Genet.* 169, 213-218 (1979).
- Hobom, G. & Lusky, M. in *Mechanistic Studies of DNA Replication and Genetic Recombination* (ed. Alberts, B.) 231-255 (Academic, New York, 1980).
- Langridge, J., Langridge, P. & Bergquist, P. L. *Analyt. Biochem.* 103, 264-271 (1980).
- Hobom, G. *et al. Cold Spring Harb. Symp. quant. Biol.* 43, 165-178 (1979).
- Schoner, B. & Kahn, M. *Gene* 14, 165-174 (1981).
- Eugler, J. A. & van Bree, M. P. *Gene* 14, 155-163 (1981).

Molecular structure of a new family of ribonucleases

Y. Mauguen*, Robert W. Hartley†, E. J. Dodson‡, G. G. Dodson‡, G. Bricogne§, Cyrus Chothia§ & A. Jack§

* Centre Pharmaceutique et LURE, Université Paris-Sud, 92290 Chateaufort-Malabry, France

† Laboratory of Nutrition and Endocrinology, National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases, Bethesda, Maryland 20205, USA

‡ Department of Chemistry, University of York, Heslington, York YO1 5DD, UK

§MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Bacillus amyloliquefaciens produces a ribonuclease¹ (barnase), the function of which is probably the digestion of external RNA: it is excreted by the bacillus and within the cell its action is inhibited by a protein of ~89 residues to which it binds with high affinity². Determination³ of the amino acid sequence of the ribonuclease, which is a monomer of molecular weight 12,382 consisting of 110 residues, has revealed its homology with other prokaryotic and eukaryotic ribonucleases⁴⁻⁶. We have now determined the atomic structure of barnase by X-ray crystallographic analysis. We report that its structure includes a large central β -pleated sheet and two α -helices. The arrangement of these secondary structures is different from that found in bovine pancreatic ribonuclease.

Crystals of barnase have the space group P3₂ and the hexagonal unit cell dimensions: $a = b = 59.0 \text{ \AA}$, $c = 81.6 \text{ \AA}$. There are three molecules in the asymmetric unit. X-ray diffrac-

|| Present address: Department of Biochemistry, Columbia University College of Physicians and Surgeons, 630 West 168 Street, New York, New York 10032, USA.

† Deceased.

Table 1 X-ray data and heavy atom refinement for barnase

Crystal	Data agreement F (R_M)	Heavy atom parameter agreement			Heavy atom reaction sites (no. in asymmetric unit)	Heavy atom contributions and errors				
		R_1	R_2	R_3			0.00	$4 \sin^2 \theta / \lambda^2$	0.08	0.12 0.16
Native	0.04									
	0.03	0.43	0.18	0.09*	His 102 (3)	$\langle FH \rangle$	262	207	169	136 110
Gold I	0.04	0.44	0.19	0.08*		$\langle EH \rangle$	62	62	79	66 50
						$\langle FH \rangle$	—	—	—	148 122
Gold II	0.04					$\langle EH \rangle$	—	—	—	77 65
						$\langle FH \rangle$	—	134	100	72 54
Iodide	0.05		0.15	0.12†	Tyr 13 (6)	$\langle EH \rangle$	—	134	126	98 80
					Tyr 17 (2)					

$$R_1 = \frac{\sum ||FHLE| - |FH_{calc}||}{\sum |FHLE|}; \quad R_2 = \frac{\sum ||FPH| - |FP||}{\sum |FPH|}; \quad R_3 = \frac{\sum ||FPH| - |FPH_{calc}||}{\sum |FPH|}$$

where $FHLE$ is the experimental value of the heavy atom contribution, FH_{calc} is the calculated heavy atom contribution⁷, $|FP|$ is the observed structure amplitude of the protein, $|FPH|$ the observed structure amplitude of the derivative, and $|FPH_{calc}|$ the derivative structure amplitude constructed by combining the calculated heavy atom contribution and the protein structure factor. $\langle FH \rangle$, r.m.s. value of FH_{calc} in the appropriate range of $4 \sin^2 \theta / \lambda^2$; $\langle EH \rangle$, r.m.s. lack of closure.

* $|FPH_{calc}|$ was constructed using the symmetry-averaged phases.

† $|FPH_{calc}|$ was constructed using the isomorphous phases defined by the gold derivative.

tion data were collected to 2.5 Å spacing on a Hilger and Watts automatic 4-circle diffractometer using the native enzyme and one derivative, which was prepared by soaking the crystals in gold cyanide. This derivative crystal diffracted well and the anomalous scattering effects produced by the gold were recorded. Owing to differences in the gold substitution, the data for the two crystals used in the case of the gold derivative were treated separately throughout. A second heavy atom derivative was prepared by iodination of the crystals and a 2.2 Å data set collected photographically at LURE using synchrotron radiation selected at $\lambda = 1.40$ Å. The anomalous scattering in this data set was not measured.

Three gold atom positions were identified from the difference Patterson function with coefficients $(|F_P| - |F_{PAu}|)^2$ and their parameters were refined by least-squares minimization to the

experimental heavy atom contribution (F_{HLE})⁷. Iodine positions were determined from the difference Fourier map (coefficients $|F_P| - |F_{PI}|$) phased by the gold derivative⁷. A phase set was calculated from the two derivatives using Blow and Crick's combination procedure⁸. Details of the heavy atom refinement and phase calculations are given in Table 1.

We were unable to interpret the electron density calculated with the combined isomorphous phases although there were regions in which the protein's main-chain and side-chain structure could be traced in detail. There was also an encouraging similarity in the electron density of each of the three independent molecules. The non-crystallographic symmetry relations between the molecules were first determined from the gold positions and then refined by correlating the electron densities within spheres centred approximately on each molecule, using

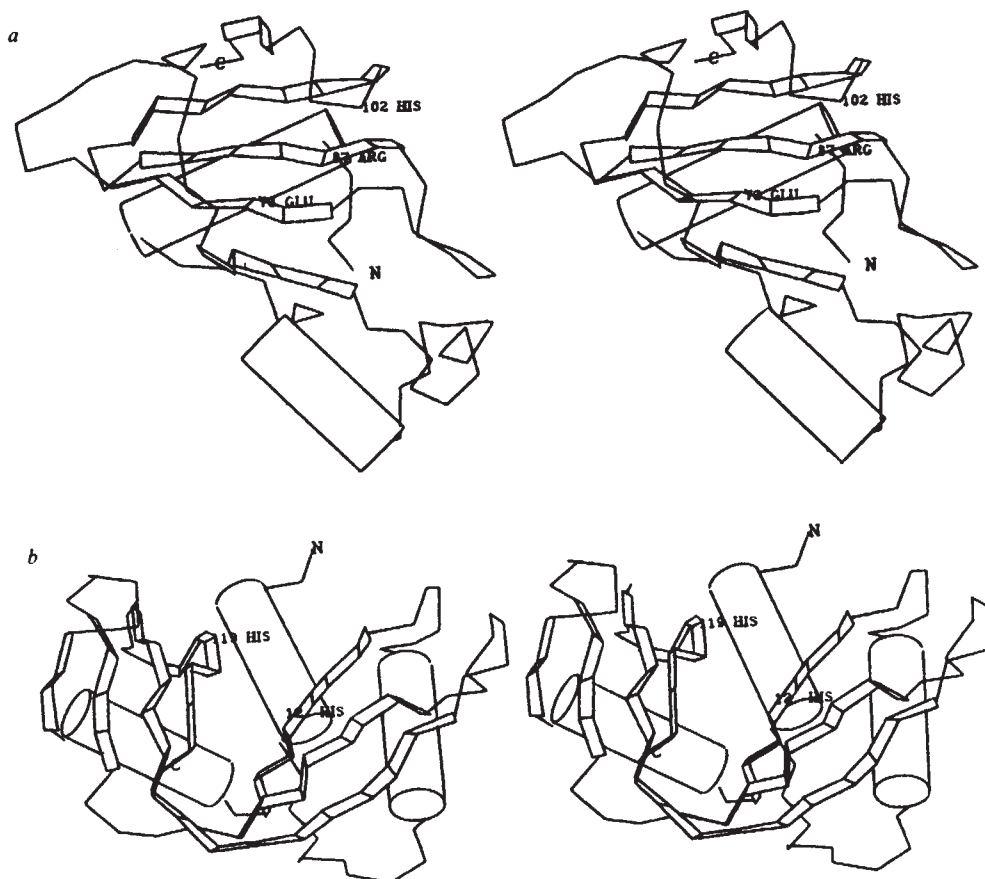


Fig. 1 Stereo pairs of schematic drawings of *a*, barnase and *b*, bovine pancreatic ribonuclease A. Each vertex represents the position of a $C\alpha$ atom; α -helices are shown as cylinders and the strands of the β -sheet as ribbons. The connecting loops are represented by lines. The active site of barnase includes Glu 73, Arg 87 and His 102 and is shown in more detail in Fig. 3. The active site of bovine pancreatic ribonuclease A includes His 12 and His 119. This figure was drawn from the atomic coordinates using the computer program of Lesk and Hardman²⁰.

a least-squares superposition method⁹. The mean of the correlation coefficients between the different molecules was 0.60.

The molecular envelope was defined from the isomorphous map, in which the electron density was averaged using the refined non-crystallographic symmetry operations. The map was sectioned in two perpendicular directions to ensure accurate tracing of the envelope. This was important to avoid the need for recalculating the envelope during the refinement calculations.

The electron density map produced with the symmetry-averaged phases^{9,10} was readily interpreted. All but two residues have sensible density, which in some cases is very well defined, and their coordinates are now being prepared for refinement calculations.

The polypeptide chain of barnase is shown in Fig. 1a. The secondary structure of the enzyme consists of a central, five-stranded, twisted antiparallel β -sheet (residues 50–55, 70–75, 85–91, 94–101 and 106–108) and two α -helices (residues 6–18 and 26–34). The first helix packs against the face of the β -sheet in the normal manner¹¹ and the second helix packs against its edge. Bovine pancreatic ribonuclease also consists of a large central β -sheet with adjacent α -helices^{12,13} but it differs from barnase in that its β -sheet is not only twisted but also bent in the middle, giving it a V shape (Fig. 1b). In addition, the residues that form the active site in the two molecules originate in parts of the structures that are not analogous (see below) and give different base specificities: barnase specific for purines¹⁴ and pancreatic ribonuclease for pyrimidines¹².

Extracellular ribonucleases from *Bacillus intermedius*, *Streptomyces erythraeus* and from the fungi *Aspergillus oryzae* and *Ustilago spheerogena*, have amino acid sequences that are homologous to that of barnase^{4–6}. Figure 2 shows that residues 30–110 of barnase are between 82 and 20% homologous with the four other sequences. There are five positions at which the residues are invariant; in barnase these positions are occupied by Leu 42, Glu 73, Ser 80, Arg 87 and His 102 (Fig. 2). The residues at three of these positions have been shown to be involved in the ribonuclease activity of the proteins: the His in barnase and ribonucleases St and T₁, the Arg in ribonucleases St and T₁, and the Glu in ribonuclease T₁ (refs 6, 15, 16).

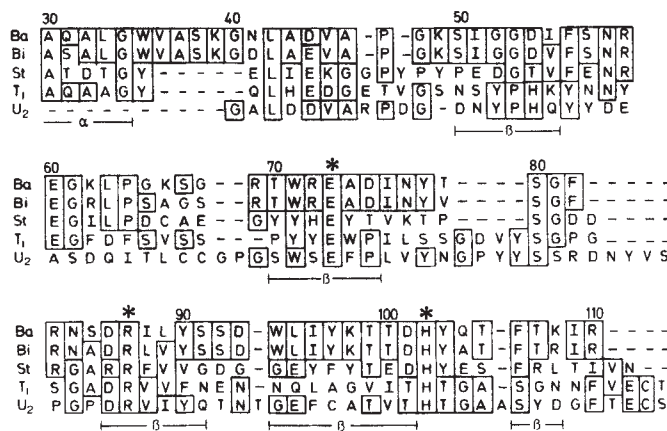


Fig. 2 An alignment of the amino acid sequences of the extracellular ribonucleases from *B. amyloliquefaciens*³ (Ba); *B. intermedius*⁵ (Bi); *S. erythraeus*⁶ (St); *A. oryzae*²¹ (T₁); and *U. spheerogena*²² (U₂). The numbering used here is that of barnase so that the positions of mutations, insertions and deletions can be seen in Fig. 1a, 3. Residues that are identical in two or more sequences are boxed and homologous residues that are part of the active site (see text) are indicated by asterisks. The regions that form secondary structures in barnase are indicated by 'α' for an α -helix and by 'β' for the strand in the β -sheet. The alignment shown here has been revised slightly from that given in ref. 4 in the view of the barnase structure. The residues that form the first part of the five sequences show little overall homology and large deletions. It has been shown that the alignment of such sequences derived by maximizing the match of the few similar or identical residues can give results that are structurally incorrect²³.

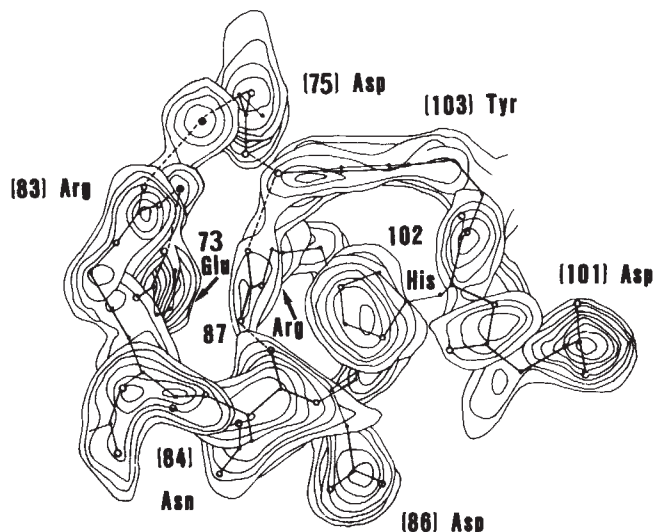


Fig. 3 A composite view of the electron densities at the active site, with the protein atomic coordinates superimposed. Two water molecules are indicated by solid circles.

Inspection of the barnase structure shows that residues Glu 73, Arg 87 and His 102 cluster in one region of the molecules, therefore this region must form the active site (see Figs 1a, 3).

The catalytic role of histidine 102 has been confirmed by the finding that gold cyanide, which reacts with imidazole N_D in the crystal, suppresses the enzyme's activity. Removal of the gold cyanide by dialysis restores fully the catalytic activity of the enzyme (R.W.H., unpublished results).

In barnase, the active-site residues are in a shallow depression and come solely from the C-terminal half of the molecule. Their chemical and structural organization differs from that at the active site of bovine pancreatic ribonuclease A, where the principal catalytic residues are two histidines, residues 12 and 119, which lie in a deep cleft (Fig. 1a, b).

Work is in progress to elucidate the catalytic mechanism of barnase and to provide a structural interpretation of previous experiments on its folding¹⁶. X-ray analyses of the ribonucleases T₁ and St are in progress elsewhere^{17–19} and comparison of their structures with that of barnase should allow an understanding of how this family of ribonucleases evolved.

We acknowledge with gratitude the use of the synchrotron radiation facilities at LURE and the help given by Peter Moody, Federico Giordano and Christopher Hill during the analysis. *Note added in proof:* Since submitting this paper we have learnt of the structure determination at 2.5 Å resolution of ribonuclease St (Yamamoto, Y. *et al.* 9th Symposium on Nucleic Acid Chemistry, Tokyo, October 1981) and of ribonuclease T (Heinemann, U. and Saenger, W., personal communication).

Received 18 January; accepted 19 March 1982.

1. Nishimura, S. & Nomura, M. *Biochim. biophys. Acta* **30**, 430–433 (1958).
2. Hartley, R. W. & Smeaton, J. R. *J. biol. Chem.* **248**, 5624–5626 (1973).
3. Hartley, R. W. & Barker, E. A. *Nature new Biol.* **235**, 15–16 (1972).
4. Hartley, R. W. *J. molec. Evol.* **15**, 355–358 (1980).
5. Aphanasenko, G. A., Dudkin, S. M., Kaminiv, L. B., Leshchikaya, I. B., Leshchikaya, I. B. & Severin, E. S. *FEBS Lett.* **97**, 77–80 (1980).
6. Yoshida, N., Sasaki, A., Rashid, M. A. & Otsuka, M. *FEBS Lett.* **64**, 122–125 (1976).
7. Dodson, E. J. *Crystallographic Computing Techniques*, 260–268 (ed. Ahmed, F. R.) (Munksgaard, Copenhagen, 1975).
8. Blow, D. M. & Crick, F. H. *Acta crystallogr.* **12**, 794–802 (1959).
9. Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).
10. Bricogne, G. *Acta crystallogr.* **A30**, 395–401 (1974).
11. Janin, J. & Chothia, C. *J. molec. Biol.* **143**, 95–128 (1980).
12. Richards, F. M. & Wyckoff, H. W. *Enzymes*, 647–806 (1971).
13. Carlisle, H. C., Palmer, R. A., Mazumdar, K. S., Gorinsky, B. A. & Yeates, D. G. R. *J. molec. Biol.* **85**, 1–18 (1974).
14. Rushizky, G. W., Greco, A. E., Hartley, R. W. & Sober, H. *Biochemistry* **2**, 787–793 (1963).
15. Takahashi, K. *J. Biochem., Tokyo* **80**, 1267–1275 (1976).
16. Hartley, R. W. *J. Biol. Chem.* **252**, 3252–3254 (1977).
17. Martin, P. D., Tulinsky, A. & Walz, F. G. Jr *J. molec. Biol.* **136**, 95–97 (1980).
18. Heinemann, U. *et al. Eur. J. Biochem.* **109**, 109–114 (1980).
19. Yamamoto, Y. *et al. J. molec. Biol.* **145**, 285–287 (1981).
20. Lesk, A. M. & Hardman, K. *Science* (in press).
21. Takahashi, K. *J. biol. chem.* **240**, 4117–4119 (1965).
22. Sato, S. & Uchida, T. *Biochemistry* **145**, 353–360 (1975).
23. Delbaere, L. T. J., Brayer, G. D. & James, M. N. G. *Nature* **279**, 165–168 (1979).

BOOK REVIEWS

From Kharkov to Tel Aviv

John M. Charap

IF YOU look up *refusenik* in a Russian dictionary, you will not find it there. After all, it is not an official Russian word, but has been improvised to describe someone who has applied to emigrate from the USSR to Israel, and whose application for a visa has been refused. What the word does not convey is the anguish and the courage of those who wait, often for years, trapped in a situation in which as individuals they face the power of a state at best indifferent, frequently hostile.

Mark Yakovlevitch Azbel is probably best known to many readers of this journal for his work on cyclotron resonance in metals. He learnt his physics at "the cradle of Soviet physics", Kharkov State University, where Lev Landau had established a world-famous research institute. From Kharkov he went to Chernogolovka, to the Landau Institute for Theoretical Physics, where he continued to rise to a position of prominence and international repute. Yet in December 1972 he took the fateful step of applying for a visa to emigrate to Israel, and became a *refusenik*. It was not until July 1977 that he was finally permitted to leave the USSR.

This remarkable book tells how it was that, notwithstanding his success and prospects for continuing advancement as a Soviet physicist, he became convinced that he could not remain in the Soviet Union. And it tells of the hazards and torments which dogged his steps from Kharkov to Tel Aviv. Azbel has written a moving and sincere autobiography. Any such work from the Soviet Union must be of interest, because of the light it sheds, sometimes no less illuminating for being oblique, on life in that country. For readers of *Nature* it will be of particular interest because of the view it gives of Soviet science, and indeed of some individual Soviet scientists. But to my mind the real importance of this book lies in its exposition, from the stand-point of one who experienced and suffered from it, of the corrupting pervasion of anti-Semitism in the USSR.

Azbel's parents were both physicians, who graduated in Kharkov in the terrible days of 1931 when famine stalked the land, the man-made famine which followed forced collectivization, and Mark was born the following year. As a four-year-old, left unsupervised whilst his parents worked six days a week, Mark first found out that he was different, that he was a Jew. Lenin had tried to eliminate the anti-Semitism which was so deeply rooted in the USSR — nowhere more than in the Ukraine. Stalin, as first Commissar for Nationalities, had made it in a certain sense impossible to be a

Refusenik: Trapped in the Soviet Union.

By Mark Ya Azbel. Pp.513. UK ISBN 0-241-10633-8; US ISBN 0-395-30226-9. (Hamish Hamilton/Houghton Mifflin: 1982.) £9.95, \$17.95.

Russian Jew, or a Ukrainian Jew. To be a Jew in the Soviet Union is not a matter of choice or conviction; it is rather a question of nationality, and if it says "Jew" on an internal passport, it cannot say "Russian" or "Ukrainian". From his first day at nursery school when his teacher greeted him with "Another zhid! Where do they all come from?" to his last few hours in Moscow when the cold insulting indifference of the officials at the airport made his departure a nightmarish farce, Azbel was time and again reminded of his difference, of his Jewishness.

Of course there were times when this awareness could recede into the background. But with dramatic force public affairs would make him, along with the whole of the Jewish community, aware of the precarious — even dangerous — status of the Jews. As a student he felt the tension which prevailed throughout the country during Stalin's last years.

The Lysenko catastrophe had repercussions all over the Soviet Union affecting biologists in even the smallest, most remote institutions, obscure little places of minor importance. Now the laws against Cosmopolitanism were threatening artists, composers, writers. Russians were not wholly exempt from these attacks; but nine out of ten of the criminals were Jews.

Confronted by the menace and threat of disappearance, arrest by the KGB or worse, Azbel asked himself: how could one survive in this atmosphere? The question was not about physical survival, but moral survival. There were several possibilities: concentration on personal life; devotion to a struggle with the system; escape into pure

science. But

I could not make any of these choices. I felt that my own way of life was to be in science, but as to friends: it seemed to me that it would be crippling, demeaning, to choose my friends on the basis of which friendships would be safe and which would not. To confine my loyalties to science, and to my fellow-scientists, for reasons of security; to cultivate indifference to everything and everyone else — I knew I couldn't do that. In a word, I simply made up my mind at that time that I would try to be a decent human being.

Stalin died, and the terrible threat implicit in the "Doctors' plot" receded. Azbel graduated, brilliantly, but not without politically motivated machinations against him which it required the intervention of Ilya Lifschitz to deflect — and not without his first encounter with the KGB, perhaps an attempt to obtain a hold on him for future exploitation. The heavy atmosphere lightened with Khrushchev's short-lived thaw. Azbel married and moved to Moscow.

Also now in Moscow was Alexander Voronel, a close friend from student days in Kharkov and now a director of a laboratory in Dubna: they had much in common, and their mutual friendship and support for Andrei Sinyavsky, and especially Yuli Daniel, brought them both into the hands of the KGB when these two authors were put on trial in 1965. This brush with the authorities was more significant than the earlier one, and Azbel was lucky to avoid losing his post at the university.

The shifting attitudes within the Soviet Union during the late 1960s, responsive to events such as the Six-Day War and the Czech uprising, meant that when in 1970 news first began to filter through that some Soviet Jews had publicly expressed a wish to emigrate to Israel, Azbel was amongst those who began to recognize that to leave the Soviet Union was a real possibility. His

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

The start of active protest: four of the seven scientists who took part in the 15-day hunger strike, June 1973, in reaction to the refusal of the Soviet authorities to allow them exit visas. From left to right: Dan Roginsky, Alexander Voronel, Mark Azbel and Victor Brailovsky.

resolve was stiffened by the trial in Leningrad of Dymshitz and Kuznetsov and the demonstrations in the Visa Department offices in Moscow in 1971.

And so it was that in December 1972 he went to the Visa Department. He was not the first distinguished scientist to do so: that was probably Alexander Lerner who had applied a year earlier. Lerner is still a *refusenik* more than ten years later. Nor was he the last. Recent information suggests there are about 100 *refuseniks* with a scientific doctorate (roughly equivalent to DSc) in Moscow alone, another 200 with the degree of Candidate (roughly PhD). Voronel applied shortly before Azbel, and was allowed to leave at the end of 1974.

It must be understood that to apply to emigrate from the USSR is an act of the highest significance. It leads almost inevitably to social ostracism. And in most cases, certainly for scientists, it leads to dismissal from one's job. For a scientist this in turn means more than loss of livelihood (which can in turn lead to prosecution for "parasitism"); it means deprivation from contact with colleagues, from access to libraries and laboratories. It means "exclusion from science". To counter that exclusion, to provide mutual support and scientific sustenance, Voronel organized a seminar every Sunday noon at his apartment. These Sunday seminars continue to this day, despite a succession of attempts to disrupt or prevent them. They have multiplied: similar meetings are held in other Soviet cities, and in disciplines other than physics. The participants have received Western visitors in this way (a recent session learnt about Feigenbaum's work on frequency doubling and the transition to chaos): indeed they have held impressive international meetings. All of this has been under the watchful eyes of the KGB, and has been scrupulously within the law.

The legality (and to Western readers perhaps the normality) of private gatherings of scientists to discuss science has not prevented harassment, and official attempts to disrupt their proceedings. These reached a climax when Voronel, Azbel and Victor Brailovsky decided in 1974 to hold an international scientific symposium, and again three years later when another international meeting, on collective phenomena in physics, was held to celebrate the fifth anniversary of the seminars. Both meetings were attended by numerous distinguished visitors from abroad. Both were heralded by oppressive KGB surveillance, questionings and indeed arrests — Azbel was himself imprisoned in 1974. After Azbel's emigration, Victor Brailovsky took over as chairman of the Sunday seminars from Azbel, who had succeeded Voronel. Brailovsky is now living in exile in Beineu, Kazakhstan.

Azbel's book speaks eloquently of the courage and resourcefulness of those scientists who persist in their legitimate

aspirations to emigrate and of their motivation. It answers some but not all of the questions which must be raised about the reasons behind the Soviet policies towards them. It underlines their vital need for support from the international scientific community, which I believe to be an obligation upon us as scientists. The pursuit of science gives privileges and brings responsibilities: and it is not possible to separate those responsibilities from the wider relations between the pursuit of

science and human affairs. This book should be read by all those who profess an interest in international scientific activities, and I would hope that would mean all scientists. □

John Charap is Head of the Physics Department at Queen Mary College, University of London. He attended the Moscow Sunday seminar in 1978, and is Chairman of the European Physical Society Advisory Committee on Scientific Freedom.

The quantum theory of space and time

C.J. Isham

Quantum Fields in Curved Space. By N.D. Birrell and P.C.W. Davies. Pp.340. ISBN 0-521-23385-2. (Cambridge University Press: 1982.) £27.50, \$49.50.

THE problem of quantizing the gravitational field commands the attention of an increasing number of theoretical physicists who are attracted by the importance of the subject to studies of the early Universe, gravitational collapse, grand unification of all fundamental forces, and the Planck length structure of space and time. In spite of much effort a successful resolution remains elusive, and it seems natural to tackle first the simpler problem of the quantization of a field propagating in a curved, but fixed, background spacetime. It had been known for many years that a time dependent metric would transmute part of its energy through the medium of quantum particle production, but the subject became of widespread interest only after Hawking's announcement in 1974 of the production of particle pairs by the event horizon of a black hole.

The subsequent effort devoted to quantum field theory in a curved space has spawned a considerable literature of research papers, conference proceedings and review articles, but the text of Birrell and Davies is the first attempt to present the subject in a more permanent form. I can think of at least three quite different types of book that might be published under this title with the contents reflecting the individual approaches that have been developed by different researchers. The present authors have made a determined and commendable effort to cite most of the important papers that have appeared, but the final treatment sharply reflects their own personal contributions and methodology. Quantum field theory is introduced with heuristic methods involving mode function expansions; there is no discussion of the mathematical problems associated with such sums or of their resolution via the operator algebra techniques that form the backbone of most rigorous investigations into quantum fields with external sources. This simpler approach does however have

the asset of rendering the book accessible to a postgraduate student who has attended any standard first course on quantum field theory, and the assumed knowledge of general relativity is no more demanding.

A significant application of the subject is to the physics of the early Universe; for example it has been suggested frequently that an initial anisotropy could be removed by the quantum production of particle pairs. Birrell and Davies are especially interested in this problem and provide a comprehensive discussion of quantum field theory in various cosmological backgrounds. Of particular value here is the careful account of the resolution, using model particle detectors, of the conceptual problems that arise concerning the operational status of a quantum particle.

Almost all work in quantum field theory is plagued by the infamous ultraviolet divergences and the addition of a curved background tends, if anything, to make matters worse. A particularly disturbing feature is the appearance of infinities in the matrix elements of the energy momentum tensor which, in a semiclassical limit, forms the source of the gravitational field. The regularization and renormalization of these divergences is a subtle problem and is discussed at length with most of the established techniques being given a respectable hearing; applications include an authoritative account of the Hawking effect and of the ensuing use of the stress tensor to isolate the origin of the particle pairs.

Professor Davies has rightfully earned a high reputation as an author of informative and well written semi-popular science books and this research monograph is presented in a similarly clear and attractive style. Quantum gravity is an active subject and it would be difficult to predict that all sections of the text will seem equally relevant in ten years time; nevertheless, the authors are to be congratulated on producing a timely work that should help to stimulate interest in this fascinating branch of theoretical physics. □

C.J. Isham is Reader in Theoretical Physics at Imperial College, University of London.

Mycology: a successor to Cochrane

John D. Weete

Fungal Physiology. By David H. Griffin. Pp.383. ISBN 0-471-05748-7. (Wiley: 1981.) £25.45, \$43.25.

DURING the past few years several excellent, advanced-level monographs have appeared on aspects of the biochemistry and physiology of fungi, reflecting some 20 years of continual advance in our understanding of these subjects. Yet there has been no successor to Cochrane's general introductory text, *Physiology of Fungi*, published by John Wiley in 1958, and long out of date. Professor Griffin has now provided such a text with *Fungal Physiology*. His coverage emphasizes those physiological processes typically associated with fungi, and stresses the experimental rather than historical aspects of the subject.

The opening three chapters, reviewing the fungal thallus, and the chemical composition and structural organization of fungus cells, are followed by an outline of the major processes occurring within fungal cells. On balance I believe more emphasis could have been given in this latter chapter to metabolism and the treatment strengthened by greater coverage of the relationships between metabolic change, growth and reproduction. The author then turns to various aspects of vegetative growth, and in the next three chapters considers the uptake of nutrients, the chemical requirements for growth and the relation of the physical environment to growth. Chapters 9 to 12 deal with sporogenesis (environmental factors, biochemistry), dormancy and germination, and the role of hormones in sexual reproduction. Here, Professor Griffin has chosen *Dictyostelium discoideum* rather than a yeast or filamentous fungus to illustrate the biochemistry of spore formation. He provides ample evidence indicating that fungi of different classes seem to have developed different hormonal systems for regulating sexual reproduction. The chapter on syngamy (Chapter 12) is organized on the basis of fungal groups and the specific hormones associated with each group are discussed.

Understanding specific biochemical processes can be aided by the use of inhibitors; in this context, inclusion of a chapter devoted to fungicides is noteworthy because of the significance of many of these compounds in the agrochemical and biomedical fields as chemical agents for controlling pathogenic fungi. The final chapter, "Fungal Attack Mechanisms", is concerned with fungal toxins, exoenzymes, fungal responses to environmental stimuli and fungal interactions (non-pathogenic) with other living systems. Fungal genetics *per se* is not covered, but genetic studies are discussed where they are relevant to other topics, particularly fungal symbiotic

relationships. The importance of using mutants in mycological research is implied by the citation of examples throughout the book.

The book is well organized; each chapter begins with an introductory statement which puts the subject under consideration into perspective, ends with a brief summary, and is appropriately referenced. Topic discussions are well illustrated with light and electron micrographs, tables and, where pertinent, graphs and chemical structures. *Fungal Physiology* is suitable as a text for a general course serving advanced undergraduate and graduate students, and also provides a handy reference work for researchers seeking general background information. The appearance of this book is both significant and timely; I highly recommend it. □

John Weete is a Professor in the Department of Botany, Plant Pathology and Microbiology at Auburn University, Alabama.

View of laser fusion

David W. Forslund

Physics of Laser Driven Plasmas. By Heinrich Hora. Pp.317. ISBN 0-471-078880-8. (Wiley: 1981.) £28.95, \$49.15.

THE field of laser fusion and the physics of laser driven plasmas has developed rapidly in the past decade and it has proven difficult for research workers to keep up with progress. Particularly lacking have been good reviews and good textbooks which would help introduce new people to this extremely complex subject. In this recent attempt to meet such a need, Professor Hora has sought to derive the basic physics of the interaction of intense laser light with matter from first principles. He attempts to lead the reader through the derivation of the fundamental plasma physics, kinetic theory, hydrodynamics and an expression for the nonlinear force arising from intense laser light, and follows this with discussion of the numerous applications of the nonlinear force from its role in parametric instabilities to its use in the implosion of pellets.

The book suffers from a number of flaws, however, which prevent it from being useful either to the student or as a handbook for the research worker. First, the discussions of most of the phenomena are unduly complex, confused and full of errors, making the text extremely difficult to read. As a typical example, the derivation of the plasma frequency and Debye length on pages 22-25 introduces

JUST
PUBLISHED

The World Environment 1972-1982

A report by the United Nations Environment Programme. Edited by Martin W. Holdgate, Mohammed Kassas and Gilbert F. White. Project Co-ordinator: Essam El-Hinnawi.

Prepared by a team of the world's most eminent environmental scientists, this study reviews the changes in the world environment since 1972, evaluates the significance of these developments and provides the foundation for international and national action in the years to come.

The book includes sections on: Atmosphere; Marine Environment; Inland Waters; Lithosphere; Terrestrial Biota; Agriculture and Forestry; Population; Human Settlements; Health; Industry; Energy; Transport; Tourism; Education and Public Understanding; Peace and Security;

xxxii + 637 pages, 117 tables, 69 figures
HARDBACK 0907567134 £50 \$95.00
SOFTCOVER 0907567142 £25 \$45.00

Global Environmental Issues

Edited by Essam El-Hinnawi and Manzur H. Hashmi.

Based on UNEP's annual state of the environment reports, concise overviews are given on major emerging environmental issues confronted during the decade. Nine areas are examined in-depth and highlighted with statistical data and illustrations.

viii + 236 pages, 106 illustrations (tables, figures, photographs).
HARDBACK 0907567118 £20 \$40.00
SOFTCOVER 0907567126 £15 \$28.50

Development without Destruction: Evolving Environmental Perceptions

By Mostafa Kamal Tolba.

A selection of Dr. Tolba's writings that provides an invaluable insight into the evolving international perceptions of major environmental issues and the important policy factors involved in planning for future action. Dr Tolba is Executive Director of the UN Environment Programme.

ix + 197 pages.
HARDBACK 0907567223 £18 \$31
SOFTCOVER 0907567231 £12 \$22

These books are part of Tycooly's Natural Resources and the Environment Series.

Tycooly
ti
International

**TYCOOLY
INTERNATIONAL
PUBLISHING
LIMITED**

6 Crofton Terrace, Dun Laoghaire,
Co. Dublin, Ireland.
Tel: (+ 353-1) 800245/6.
Telex: 30547 SHCN EI

the unclear and irrelevant ideas of "cell distance", "refractive index" and "electrostatic explosion of ions".

Throughout the book Professor Hora seems to be forcing his own conceptions on the reader rather than trying to properly survey the field. As a consequence the book gives a very unbalanced view. For example, only one paragraph is devoted to the production of high energy electrons, which continues to be such a major difficulty in laser fusion. Consistent with this neglect, the major problem of electron transport is totally ignored.

More than one-third of the book is spent on calculating the ponderomotive force and its effects on hydrodynamics, which leads, in the conclusion of Chapter 13, to the claim that all the problems of hot electrons in experiments can be removed by utilizing "the nonlinear force". As an alternative to the usual ablative

compression approach which suffers from electron preheat, Professor Hora suggests . . . the very fast nonlinear force pusher will accelerate the plasma before the hot electrons can be generated. This has been confirmed by extensive numerical studies and can be the way out of the present difficulties of laser fusion.

Apparently, Professor Hora genuinely believes that a short pulse intense laser at the upper limit of already achieved intensities will not produce energetic electrons. It is remarkable that, although this simple target approach is at least five years old, no major experimental programme has sought to make use of it. Evidently, of the many proposals to reduce the preheat problem, Professor Hora's has been judged to be the least likely to succeed. □

David W. Forslund is a Fellow of the Los Alamos National Laboratory, New Mexico.

Seismologists look at arms control

Peter D. Marshall

Identification of Seismic Sources — Earthquake or Underground Explosion. Edited by Eystein S. Husebye and Svein Mykkeltveit. Pp. 876. ISBN 90-277-1320-0. (Reidel: 1981.) Dfl. 195, \$98.

IN 1958 a Committee of Experts met in Geneva to study the possibility of verifying a treaty to ban the testing of nuclear devices. Under the treaty, eventually signed in 1963, the USA, USSR and UK stopped all but underground testing — although underground explosions could be detected by their seismic effects, uncertainty in discriminating between explosions and the many earthquakes of equivalent magnitude meant that a treaty to ban them was politically unacceptable. This uncertainty in discrimination has occupied forensic seismologists for over 20 years. Their latest views are aired in these proceedings of a NATO-sponsored Study Institute, convened to assess progress since systematic research began with the Vela Uniform programme in the USA and Tabor Pluto in the UK.

These programmes led to instrumental improvements and to the use of digital computers to increase the accuracy of detection and hypocentre location. Many more earthquakes could then be distinguished from explosions because of their greater depth; in addition, the spectral differences between earthquake and explosion signals led to identification by their body to surface-wave ratios. These improvements opened up the possibility of teleseismic rather than the short-range monitoring available in 1958, thereby allowing a potential reduction in the number of stations required to verify a treaty. However problems continued to arise: the increasing number of explosions in both seismic and aseismic regions brought differences in transmission paths to the attention of seismologists and provided much of the impetus for modelling seismograms. This in turn highlighted the loss of information suffered by the narrow recording bands in common use, and thought and effort was successfully brought to bear on digital recording and processing in much wider bands.

By 1976 many countries had initiated seismological discrimination studies and the UN Committee for Disarmament took advantage of this wider interest to set up a multinational Ad Hoc Group of Scientific Experts to make recommendations on how a Comprehensive Test Ban Treaty (CTBT) might be monitored. Soon afterwards, CTBT negotiations were re-opened between the USA, USSR and UK; these demonstrated a need to supplement teleseismic stations with regionally sited stations in order to achieve the verification levels required.

At this point, groups in the USA and

Norway proposed a discussion of recent developments in discrimination studies under NATO sponsorship. The resulting proceedings of the Study Institute provide seismologists with a reference book of 45 papers, 20 by invited key lecturers.

The scene is set by Douglas's objective review which includes a comprehensive account of diagnostic criteria. Effective and stable diagnostic criteria depend on a knowledge of earthquake and explosive sources. Not surprisingly little is known about the earthquake source but strangely, as Rodean shows, much the same applies to nuclear explosions even after 20 years of underground testing. Any technique which defines their parameters thus becomes a significant diagnostic aid. One in which the seismic moment tensor is used to represent the source is discussed, among others, by Doornbos, clarifying what was becoming a rather confused topic and one turning out to be well worth further exploitation even though the method is bedevilled by transmission path effects. Waveform synthesis, discussed by both Harkrider and Kennett, may however one day lead to a solution of the problem of path effects. Of the seven papers on the influences of attenuation and scattering (both path effects) on the waveforms, that by Aki demonstrates the dramatic effects of even small velocity fluctuations within the lithosphere beneath a recording station.

In spite of the uncertainties, a seismologist must ultimately make a decision: earthquake or explosion? Much may depend on his decision, so the more independent observations available to him the better and Tjostheim reports on how multidimensional discrimination techniques may be applied to teleseismic data. The technique can also be applied to regional diagnostic data of the type described by Blandford. Finally, an indication is given of the kind of network which would need to be established with the appropriate instrumentation, automatic processing, data transmission facilities and data centres capable of promptly handling enormous quantities of data.

In 1958 Sir William (later Lord) Penney described seismology as a "stone-age science". The contents of this book demonstrate that this is no longer true, and the organizers of the meeting are to be congratulated on their initiative. Nonetheless problems remain. Seismic decoupling of explosions, explosion-like earthquakes, the detection of small explosions among earthquake signals and, above all, inadequate knowledge of the structure of the Earth all contribute to the difficulties of concluding a CTBT. Seismologists and physicists anxious not to rediscover the wheel will find in this volume the sources on which to build their new research programmes. □

Peter D. Marshall is a Principal Scientific Officer with the Ministry of Defence, Blacknest, Berkshire.

● A revised, paperback edition of *The Question of Animal Awareness* by Donald R. Griffin has been published by William Kaufmann Inc., price £6.20, and is available through W.H. Freeman. The original edition was reviewed in *Nature* 266, 792; 1977.

● A.I. Sabra's *Theories of Light, from Descartes to Newton*, first published by the Oldbourne Book Co. in 1967, has been re-issued by Cambridge University Press in both hardback and paperback. The book is essentially a reprint of the original edition, but some minor corrections and a new bibliography have been included. Price is hbk £20, pbk £6.95.